



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΦΟΡΤΙΩΝ ΜΕ ΜΗ ΠΑΡΕΜΒΑΤΙΚΗ ΤΕΧΝΙΚΗ

Διπλωματική Εργασία

Κωνσταντίνος Δαμάσκος

Επιβλέπων: Δημήτριος Μπαργιώτας

Βόλος 2020



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΦΟΡΤΙΩΝ ΜΕ ΜΗ ΠΑΡΕΜΒΑΤΙΚΗ ΤΕΧΝΙΚΗ

Διπλωματική Εργασία

Κωνσταντίνος Δαμάσκος

Επιβλέπων: Δημήτριος Μπαργιώτας

Βόλος 2020



UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NON-INTRUSIVE LOAD MONITORING

Diploma Thesis

Konstantinos Damaskos

Supervisor: Dimitrios Bargiotas

Volos 2020

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα εργασία αποτελεί διπλωματική εργασία στα πλαίσια του προπτυχιακού προγράμματος σπουδών του τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών. Πριν την παρουσίαση των αποτελεσμάτων της παρούσας διπλωματικής εργασίας, αισθάνομαι την υποχρέωση να ευχαριστήσω ορισμένους από τους ανθρώπους που γνώρισα, συνεργάστηκα μαζί τους και έπαιξαν πολύ σημαντικό ρόλο στην πραγματοποίησή της. Κατ' αρχήν, θα ήθελα να ευχαριστήσω την οικογένειά μου, για την υποστήριξη διότι δεν θα τα είχα καταφέρει χωρίς αυτούς. Στην συνέχεια, θα ήθελα να ευχαριστήσω τον επιβλέποντα της διπλωματικής μου εργασίας, τον Καθηγητή Δημήτριο Μπαργιώτα για την υποστήριξη, την κατανόηση και τον σεβασμό που μου έδειχνε όλο αυτόν τον καιρό. Τέλος, θα ήθελα να ευχαριστήσω από καρδιάς τους φίλους μου και συναδέλφους μου που με στήριζαν καθ' όλη τη φοιτητική μου ζωή.

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

«Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής».

Ο/Η Δηλών/ούσα

Κωνσταντίνος Δαμάσκος
Ημερομηνία

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία πραγματεύεται το θέμα της παρακολούθησης ηλεκτρικού φορτίου με μη παρεμβατική τεχνική (Non-Intrusive Load Monitoring-NILM). Οι μέθοδοι παρακολούθησης φορτίων χωρίς την χρήση έξυπνων μετρητών που παρεμβάλλονται ανάμεσα στις συσκευές και τα σημεία τροφοδοσίας τους, επιδιώκουν να προβούν στον επιμερισμό της ενέργειας με σκοπό την εκτίμηση της κατανάλωσης κάθε συσκευής από την συνολική ενέργεια της κατοικίας. Σε αυτή την εργασία παρουσιάζονται διάφορες έννοιες σχετικές με το θέμα της αποσύνθεσης της ενέργειας, αναλύονται οι τρόποι και τα μέσα επιμερισμού και εκτελούνται διάφορα πειράματα με την εργαλειοθήκη NILMtk.

ABSTRACT

This thesis deals with the topic of Non-Intrusive Load Monitoring (NILM). Load monitoring methods without the use of smart meters inserted between the devices and their power stations, seek to disaggregate the energy in order to estimate the total energy consumption of each device in the home. This paper presents various concepts related to the issue of energy decomposition, analyzes ways and means of disaggregation, and performs various experiments with the NILMtk toolkit.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ	vi
ABSTRACT	vii
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ.....	viii
ΚΕΦΑΛΑΙΟ 1.....	1
Εισαγωγή	1
1.1 Το Ενεργειακό Πρόβλημα	1
1.2 Δίκτυα Ηλεκτρικής Ενέργειας	2
1.2.1 Τα Συμβατικά Ηλεκτρικά Δίκτυα	2
1.2.2 Έξυπνα Δίκτυα	4
1.3 Επιτήρηση Φορτίου	6
1.3.1 Παρεμβατική επιτήρηση Φορτίου	6
1.3.2 Μη Παρεμβατική επιτήρηση Φορτίου	9
ΚΕΦΑΛΑΙΟ 2.....	11
Ο Επιμερισμός της Ενέργειας	11
2.1 Ορισμός.....	11
2.2 Η Μεθοδολογία της Μη Παρεμβατικής Παρακολούθησης Φορτίου.....	12
2.2.1 Διάκριση Συσκευών.....	13
2.2.2 Μονάδες Απόκτησης Δεδομένων	16
2.2.3 Εξαγωγή Χαρακτηριστικών	18
2.2.4 Αναγνώριση Φορτίου	20
ΚΕΦΑΛΑΙΟ 3.....	22
Μεθοδοί NILM	22
3.1 Βασικές Έννοιες.....	22

3.1.1 Μηχανική Μάθηση.....	22
3.1.2 Νευρωνικά Δίκτυα	24
3.2 Μαρκοβιανά Μοντέλα	27
3.2.1 Αλυσίδες Markov	27
3.2.2 Κρυπτομαρκοβιανά Μοντέλα.....	29
3.2.3 Παραγοντικά Κρυπτομαρκοβιανά Μοντέλα	31
3.3 Αλγόριθμοι Αποσύνθεσης	33
3.3.1 DDSC	33
3.3.2 CO	36
3.3.3 HMM	40
3.3.4 MLE	42
ΚΕΦΑΛΑΙΟ 4.....	43
Δεδομένα & Εργαλεία.....	43
4.1 Σύνολα Δεδομένων	43
4.1.1 REDD.....	43
4.1.2 UK-DALE.....	44
4.1.3 BLUED	45
4.1.4 ECO	45
4.1.5 Pecan Street	46
4.1.6 iAWE	47
4.1.7 NILM Datasets	47
4.2 NILMTK.....	50
4.2.1 Λόγοι για NILMTK	50
4.2.2 Επισκόπηση του NILMTK	51
4.2.3 Κλάσεις	55

ΚΕΦΑΛΑΙΟ 5.....	60
Πειράματα	60
5.1 Εργαλεία	60
5.2 Εκτέλεση.....	61
5.2.1 CO-FHMM.....	61
5.2.2 Hart	71
ΚΕΦΑΛΑΙΟ 6.....	75
Επιλογος	75
6.1 Δυσκολίες	75
6.2 Συμπεράσματα	75
ΒΙΒΛΙΟΓΡΑΦΙΑ	76

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 Το Ενεργειακό Πρόβλημα

Η ενέργεια είναι ένα αναπόσπαστο κομμάτι της ζωής των ανθρώπων με τεράστια επίδραση σε αυτούς και το περιβάλλον τους. Κάποιες φορές είναι δυσεύρετη και ιδιαίτερα ακριβή, γεγονός που φέρει μεγάλες επιπτώσεις που έχουν να κάνουν με την έλλειψη άμεσων ενεργειακών υπηρεσιών μέχρι και τον πληθωρισμό, την ανεργία και την μείωση της οικονομικής παραγωγής ενός κράτους. Κάποιες άλλες φορές η προσφορά της είναι τόσο μεγάλη, δυσανάλογη των διαθέσιμων πόρων για την παράγωγή της, με αποτέλεσμα να ελλοχεύει ο κίνδυνος της καταστροφής του φυσικού περιβάλλοντος. Τα μεγέθη, οι ρυθμοί ανάπτυξης, η γεωγραφική κατανομή του ανθρώπινου πληθυσμού καθορίζουν τον βαθμό της ζήτησης και τις απαιτήσεις για ενέργεια αλλά και την επιλογή των μέσων που θα επιστρατευθούν για να καλύψουν επαρκώς τις απαιτήσεις αυτές. Αντίθετα, η διαθεσιμότητα, το κόστος, οι επιπτώσεις και η αποτελεσματικότητα με την οποία η ενέργεια αξιοποιείται είναι και θα παραμείνουν σημαντικοί παράγοντες που θα καθορίζουν το σύνολο των ανθρώπων στους οποίους θα παρέχεται και την ποιότητα ζωής τους.

Ο τρόπος αξιοποίησης της ενέργειας από ένα κράτος, ένα εργοστάσιο, ένα μικρό κατάστημα ή ένα νοικοκυριό επηρεάζει, το καθένα σε ανάλογο βαθμό, την ζήτηση της. Ο βιομηχανικός, ο καταναλωτικός μας πολιτισμός και ο καθημερινός μας τρόπος ζωής είναι άρρηκτος συνδεδεμένος με τον τομέα της ενέργειας. Συνεπώς, είναι συνδεδεμένος και με την κατανάλωση των πόρων που είναι αναγκαίοι για την παραγωγή της απαιτούμενης ενέργειας και με τις όποιες οικονομικές και περιβαλλοντικές επιπτώσεις συνεπάγονται. Αν λάβουμε υπόψη την ιδιαίτερα ανησυχητική μείωση των αποθεμάτων υδρογονάνθρακα, την σημαντικότερη πηγή ενέργειας παγκοσμίως, εύλογο είναι όχι μόνο να στραφούμε σε νέα μέσα και τεχνολογίες παραγωγής ενέργειας, αλλά και να

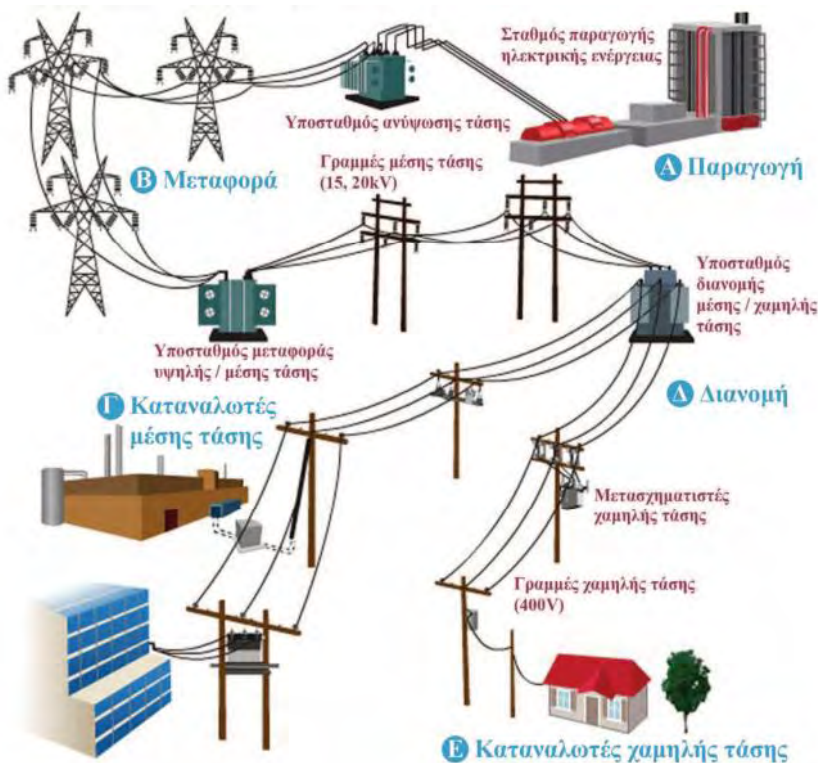
αναθεωρήσουμε τον τρόπο με τον οποίο αξιοποιούμε την ενέργεια. Όλα τα παραπάνω, αποκτούν μεγάλο βάρος στον κτιριακό τομέα που σύμφωνα με έρευνες καταλαμβάνει μεγάλο μερίδιο στην αγορά και στην κατανάλωση ενέργειας [1].

1.2 Δίκτυα Ηλεκτρικής Ενέργειας

1.2.1 Τα Συμβατικά Ηλεκτρικά Δίκτυα

Μέσο ενός ηλεκτρικού δικτύου επιτυγχάνεται η παραγωγή, η μεταφορά και η διανομή της ηλεκτρικής ενέργειας από τον παραγωγό στον καταναλωτή. Ένα συμβατικό ηλεκτρικό δίκτυο απαρτίζεται από τους σταθμούς παραγωγής της ηλεκτρικής ενέργειας, τους υποσταθμούς μεταφοράς ανύψωσης τάσης, τους υποσταθμούς για την μεταφορά υψηλής τάσης και υποβιβασμού σε μέση, τους υποσταθμούς διανομής μέσης τάσης και υποβιβασμού σε χαμηλή τάση και φυσικά από τις γραμμές διανομής υψηλής, μέσης και χαμηλής τάσης αντίστοιχα. Συγκεκριμένα, η ενέργεια παράγεται στους σταθμούς παραγωγής είτε από ανανεώσιμες πηγές όπως είναι η ηλιακή και η αιολική ενέργεια είτε από μη ανανεώσιμες όπως τα ορυκτά καύσιμα ή η πυρηνική ενέργεια. Στην συνέχεια, μεταφέρεται σε υποσταθμούς ανύψωσης τάσης, όπου μέσω μετασχηματιστών η τάση του ηλεκτρικού ρεύματος ανυψώνεται ώστε να ελαχιστοποιηθούν οι απώλειες κατά την μετέπειτα μεταφορά της λόγω του φαινομένου Joule. Έπειτα μεταφέρεται μέσω το γραμμών υψηλής τάσης στους υποσταθμούς μεταφοράς όπου μετασχηματίζεται η υψηλή τάση σε μέση για να πάρουν σειρά οι υποσταθμοί διανομής, όπου μετασχηματίζουν την μέση τάση σε χαμηλή και μεταφέρουν την ενέργεια μέσω των γραμμών μεταφοράς χαμηλής τάσης στον καταναλωτή. Η αναπαράσταση της παραπάνω διαδικασίας φαίνεται στην Εικόνα 1.1. Στο σημείο αυτό, είναι σημαντικό να προσθέσουμε ότι οι παραπάνω υποσταθμοί εκτός από τον μετασχηματισμό της τάσης χρησιμεύουν στην διασύνδεση υφιστάμενων υποσταθμών με εναλλακτικές τροφοδοσίες σε περίπτωση πιθανής αδυναμίας υποσταθμού για χρήση. Παράλληλα, οι υποσταθμοί μεταφοράς είναι εξοπλισμένοι με διακόπτες κυκλωμάτων ώστε να καθίσταται δυνατή η διακοπή παροχής ενέργειας ή η απομόνωση ενός μέρους του δικτύου σε περίπτωση σφάλματος, βλάβης, συντήρησης κτλ. Από τους

υποσταθμούς δεν λείπουν ακόμα και οι μονάδες συλλογής δεδομένων που τροφοδοτούν τα κέντρα ελέγχου της ενέργειας με τις απαραίτητες πληροφορίες.



Εικόνα 1.1: Συμβατικό δίκτυο ηλεκτρικής ενέργειας.

Η λειτουργία ενός συμβατικού δικτύου, η διαδικασία δηλαδή κατά την οποία η ενέργεια ξεκινάει από τους σταθμούς παραγωγής και καταλήγει στους καταναλωτές αναλαμβάνεται από τον Διαχειριστή Συστήματος Μεταφοράς. Επομένως, βασικός υπεύθυνος της αποδοτικής και ομαλής λειτουργίας του δικτύου είναι ο Διαχειριστής Συστήματος Μεταφοράς. Ταυτόχρονα, απαραίτητη κρίνεται η διαρκής παρακολούθηση του δικτύου μέσω κατάλληλων συστημάτων και μονάδων για την αποκατάσταση ενδεχόμενων διακοπών της παροχής της ηλεκτρικής ενέργειας αλλά και ακριβέστερης πρόβλεψης του φορτίου-ζήτησης για ισοζυγισμό της παραγωγής. Η παραπάνω λειτουργία εκτελείται από το Σύστημα Εποπτικού Ελέγχου και Συλλογής Πληροφοριών. Αναλυτικότερα το συγκεκριμένο σύστημα αναλαμβάνει την διεκπεραίωση των

παρακάτω βασικών ενεργειών: συλλογή, επεξεργασία, παρουσίαση δεδομένων και έλεγχο του συστήματος. Ένα τέτοιο σύστημα, προκειμένου να φανεί χρήσιμο συνεργάζεται πάντα με ένα άλλο σύστημα, το οποίο αξιοποιεί τα παραπάνω επεξεργασμένα δεδομένα για να βελτιώσει την διαδικασία παραγωγής και μεταφοράς της ενέργειας. Αυτό το σύστημα ονομάζεται Σύστημα Διαχείρισης Ενέργειας και μέσω διάφορων εφαρμογών συμβάλει στην αποδοτική, ασφαλή και οικονομική λειτουργία του δικτύου. Μια κλασσική λειτουργία του Συστήματος Διαχείρισης Ενέργειας είναι η καλύτερη αξιοποίηση των σταθμών παραγωγής δεδομένης της πρόβλεψης ζήτησης.

Η εξέλιξη της τεχνολογίας έχει φέρει τεράστιες αλλαγές και προόδους σε πολλούς τομείς. Ωστόσο, το δίκτυο ηλεκτρικής ενέργειας φαίνεται να έχει χάσει έδαφος σχετικά με τον εκσυγχρονισμό του. Μια μεγάλη αδυναμία του είναι η μονόπλευρη επικοινωνία και διανομή της ενέργειας, καθώς το συμβατικό δίκτυο υποστηρίζει την παραγωγή ενέργειας μόνο στους σταθμούς παραγωγής και δεν προωθεί τους καταναλωτές να συμβάλουν στη καλύτερη λειτουργία του δικτύου, είτε παράγοντας οι ίδιοι ενέργεια είτε συλλέγοντας και μοιράζοντας χρήσιμα δεδομένα. Έτσι λοιπόν, φαντάζει αδήριτη ανάγκη η τροποποίηση του συμβατικού δικτύου σε κάποιο «εξυπνότερο» όπου οι τεχνολογίες της επικοινωνίας και της πληροφορίας θα έχουν σημαντικό ρόλο.

1.2.2 Έξυπνα Δίκτυα

Έξυπνο δίκτυο ή ευφυές δίκτυο ορίζουμε ένα ηλεκτρικό δίκτυο που μπορεί να παρέχει ηλεκτρική ενέργεια με ελεγχόμενο και έξυπνο τρόπο από τις μονάδες παραγωγής στους καταναλωτές. Συγκεκριμένα, ένας από τους επίσημους ορισμούς είναι: το αναβαθμισμένο δίκτυο ηλεκτρικής ενέργειας στο οποίο έχουν προστεθεί αμφίδρομη ψηφιακή επικοινωνία μεταξύ καταναλωτή και προμηθευτή καθώς και τα έξυπνα συστήματα μέτρησης, παρακολούθησης και ελέγχου των παραμέτρων της ηλεκτρικής ενέργειας (Εικόνα 1.2). Επομένως, τα δίκτυα τέτοιου τύπου έχουν κοινό γνώρισμα την ενσωμάτωση τεχνολογιών επικοινωνιών και πληροφορικής.



Εικόνα 1.2: Σύστημα έξυπνου δικτύου.

Τα έξυπνα δίκτυα έχουν τα εξής χαρακτηριστικά γνωρίσματα:

- Ροή πληροφορίας από την πλευρά του προμηθευτή και την πλευρά του καταναλωτή σε πραγματικό χρόνο. Στο ηλεκτρικό δίκτυο παράγεται τεράστιος όγκος πληροφορίας που είναι διαθέσιμος για αξιοποίηση από τους διαχειριστές, τους παραγωγούς, τους προμηθευτές και τους καταναλωτές. Οι παραγωγοί και οι διαχειριστές λαμβάνουν χρήσιμα δεδομένα για την ακριβέστερη πρόβλεψη της ζήτησης ενώ οι καταναλωτές αποκτούν εποπτεία στις ποσότητες ενέργειας που καταναλώνουν, ελέγχοντας συνεπώς το κόστος λειτουργίας των συσκευών τους.
- Βέλτιστη αξιοποίηση του εξοπλισμού του συστήματος ηλεκτρικού δικτύου. Η αποδοτικότερη αξιοποίηση των μονάδων και των μέσων του δικτύου σε συνδυασμό με την συνεισφορά των καταναλωτών στην παραγωγή ενέργειας μεταθέτει την μελέτη και την κατασκευή νέων μέσων για την παραγωγή και την διανομή της ενέργειας.
- Δυνατότητα αυτόματης αποκατάστασης βλαβών. Η χρήση κατάλληλων αισθητήρων σε καίρια σημεία του δικτύου και η ενσωμάτωση έξυπνων συσκευών περιορίζουν το ανθρώπινο παράγοντα όσον αφορά τον έλεγχο και την συντήρηση του δικτύου. Το δίκτυο έχει την ικανότητα να αυτοεπιδιορθώνεται και ανταποκρίνεται γρηγορότερα στα πιθανά σφάλματα.

- Οι ανανεώσιμες πηγές ενέργειας έχουν μεγαλύτερη συμμετοχή στην παραγωγή ενέργειας. Η φύση της λειτουργίας του δικτύου δίνει την δυνατότητα μεγαλύτερης κάλυψης της ανάγκης για ενέργειας από τις ΑΠΕ [2].

1.3 Επιτήρηση Φορτίου

Όπως είδαμε, χαρακτηριστικό γνώρισμα των έξυπνων δικτύων είναι η αμφίδρομη ροή πληροφορίας μεταξύ του καταναλωτή και του προμηθευτή. Τα ενεργειακά δεδομένα μπορούν να αξιοποιηθούν τόσο από την πλευρά του καταναλωτή όσο και από την πλευρά του καταναλωτή. Οι καταναλωτές έχοντας γνώση την καταναλισκόμενη ενέργειας των συσκευών τους μπορούν να σχεδιάσουν και να εφαρμόσουν στρατηγικές για την οικονομικότερη λειτουργία των συσκευών και οι εταιρίες να ερευνήσουν την χρήση των συσκευών με σκοπό την αποδοτικότερη λειτουργία τους στα μέτρα των αναγκών των καταναλωτών. Ωστόσο, ένα τέτοιο εγχείρημα προϋποθέτει τις κατάλληλες τεχνολογίες, υποδομές και τον ανάλογο εξοπλισμό. Υπάρχουν δύο κύριες μέθοδοι επιτήρησης του φορτίου ενός ηλεκτρικού δικτύου και διακρίνονται στην παρεμβατική και μη παρεμβατική μέθοδο, έννοιες που θα αναλύσουμε στην συνέχεια.

1.3.1 Παρεμβατική επιτήρηση Φορτίου

Με την έννοια παρεμβατική επιτήρηση φορτίου (Intrusive Load Monitoring) αναφερόμαστε στην διαδικασία κατά την οποία παρακολουθούμε την ηλεκτρική συμπεριφορά ενός φορτίου μέσω μιας ειδικής συσκευής που παρεμβάλλει ανάμεσα στην συσκευή που επιθυμούμε να μετρήσουμε και το σημείου τροφοδοσίας της. Μια τέτοια συσκευή είναι ο έξυπνος μετρητής (smart meter) (Εικόνα 1.3).



Εικόνα 1.3: Το μετρικό όργανο Smart Meter.

Ο έξυπνος μετρητής είναι μία προηγμένη συσκευή ενεργειακής μέτρησης που καταγράφει την καταναλισκόμενη ενέργεια ενός ηλεκτρικού φορτίου. Οι έξυπνοι μετρητές έχουν την δυνατότητα να καταγράφουν την καταναλισκόμενη ενέργεια σε πραγματικό χρόνο, να αναγνωρίζουν την τιμή της τάσης, την γωνία φάσης και την συχνότητα. Επιπρόσθετα, υποστηρίζουν την αμφίδρομη επικοινωνία μεταξύ αυτών των καταναλωτών και των προμηθευτών ηλεκτρικής ενέργειας (Εικόνα 1.4). Ένα σύστημα με smart meters περιλαμβάνει τις συσκευές μέτρησης, κατάλληλη υποδομή για την επικοινωνία και συσκευές ελέγχου. Οι έξυπνοι μετρητές δεν αρκούνται μόνο στον παθητικό ρόλο της καταγραφής και αποστολής δεδομένων αλλά δύνανται να ελέγχουν την παροχή ρεύματος στις συσκευές που παρακολουθούν. Για παράδειγμα, ένας έξυπνος μετρητής μπορεί να προγραμματιστεί με τέτοιο τρόπο ώστε να διακόψει την τροφοδοσία σε ένα φορτίο αν η κατανάλωσή της υπερβεί κάποιο προκαθορισμένο όριο. Μπορούν να δεχθούν εντολές είτε τοπικά είτε από απόσταση και η επικοινωνία μεταξύ τους και με το δίκτυο επιτυγχάνεται είτε μέσω ενσύρματων συνδέσεων είτε μέσω ασύρματης σύνδεσης Wi-Fi ή cellular data.



Εικόνα 1.4: Η αμφίδρομη επικοινωνία μεταξύ μετρητών, καταναλωτών και εταιριών παραγωγής.

Γίνεται εύκολα αντιληπτό ότι αυτές οι επιπλέον πληροφορίες που έρχονται να προσφέρουν οι έξυπνοι μετρητές σε επίπεδο κατανάλωσης συσκευών μπορεί να αλλάξει ριζικά την κατανάλωση της ηλεκτρικής ενέργειας και επομένως να προσθέσει νέες δυνατότητες στο δίκτυο. Ο καταναλωτής έχοντας την δυνατότητα να έχει επίγνωση του ποσού της ενέργειας που καταναλώνει κάθε συσκευή του, κάθε ώρα ή ακόμα και κάθε λεπτό μπορεί να υπολογίσει με ακρίβεια την οικονομική επιβάρυνση που θα έχει. Συνεπώς, θα ελέγχει καλύτερα την χρήση των συσκευών του με γνώμονα την οικονομικότερη λειτουργία τους. Εύλογα βγαίνει το συμπέρασμα ότι οι νέοι αυτοί μετρητές είναι σε θέση να μειώσουν σημαντικά την ζήτηση. Μειωμένη ζήτηση ενέργειας μεταφράζεται σε μικρότερες απαιτήσεις φυσικών πόρων, δηλαδή σε μικρότερη καταπόνηση του ήδη επιβαρυσμένου φυσικού περιβάλλοντος.

Σε αυτό το σημείο είναι σημαντικό να επισημάνουμε ότι κάθε τεχνολογία μαζί με τα οφέλη της έχει και κάποια μειονεκτήματα. Έτσι λοιπόν και η χρήση των έξυπνων μετρητών εμφανίζει και κάποιες προσκλήσεις που καλούνται να αντιμετωπίσουν τόσο οι πάροχοι ηλεκτρικού ρεύματος όσο και οι καταναλωτές. Πριν προχωρήσουμε στην

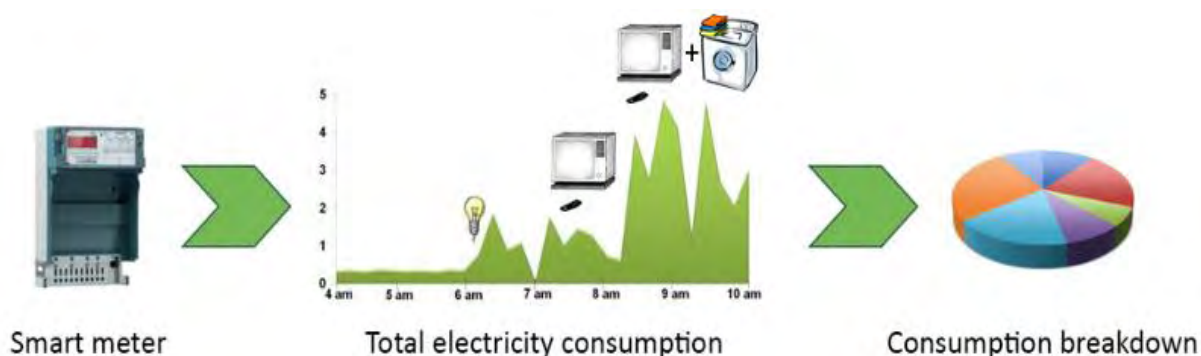
ανάλυση αυτών, είναι σημαντικό αρχικά να διευκρινίσουμε τους τρόπους χρήσης των έξυπνων μετρητών. Ο επικρατέστερος τρόπος χρήσης των συστημάτων αυτών είναι η παρεμβολή μεταξύ συσκευής και σημείου τροφοδοσίας με συσκευές-βύσματα παρακολούθησης. Στην σημερινή εποχή, σε ένα μέσο οικισμό οι συνδεδεμένες ηλεκτρικές συσκευές στο δίκτυό του είναι πολλές πράγμα που σημαίνει ότι η παρακολούθηση και η καταγραφή της ενεργειακής τους συμπεριφοράς απαιτεί την τοποθέτηση αντίστοιχου αριθμού βυσμάτων-συσκευών. Κάτι τέτοιο, είναι ιδιαίτερα κοστοβόρο και έρχεται σε σύγκρουσή με το πρωταρχικό κίνητρο του καταναλωτή που δεν είναι άλλο από την εξοικονόμηση χρημάτων. Από πλευράς προμηθευτή, η επεξεργασία και η αποθήκευση του τεράστιου όγκου δεδομένων που ενδέχεται να προκύψουν από τους μετρητές, ώστε να οδηγήσουν σε χρήσιμα συμπεράσματα, προϋποθέτει την κατάλληλη εκπαίδευση και εξειδίκευση του προσωπικού του αλλά και τον κατάλληλο εξοπλισμό, δηλαδή άλλη μια ακόμα κοστοβόρα διαδικασία. Επίσης, η διανομή δεδομένων των καταναλωτών στο δίκτυο, κλονίζουν την προστασία της ιδιωτικότητάς τους, καθώς μπορεί για κάποιους τα δεδομένα σχετικά με την λεπτομερή παρακολούθηση διάφορων συσκευών να είναι ευαίσθητες πληροφορίες. Σε όλα αυτά λοιπόν, έρχεται να προστεθεί και η δημιουργία και η συντήρηση ενός τοίχους προστασίας για την ασφάλεια των προσωπικών δεδομένων και του ελέγχου των έξυπνων μετρητών. Από την στιγμή που μία συσκευή συνδέεται στο διαδίκτυο εγκυμονεί ο κίνδυνος να παρθεί ο έλεγχος ή να καταγραφεί η δραστηριότητά της από τρίτους [3].

1.3.2 Μη Παρεμβατική επιτήρηση Φορτίου

Στην προηγούμενη ενότητα αναφέραμε τους λόγους που η παρακολούθηση φορτίου φαίνεται να είναι η νέα τάση της εποχής και πώς οδηγεί στην μείωση της κατανάλωσης ενέργειας με ότι αυτό συνεπάγεται. Ασχοληθήκαμε με την παρεμβατική μέθοδο επιτήρησης φορτίου και συγκεκριμένα με τους έξυπνους μετρητές. Αν και η μέθοδος παρεμβατικής τεχνικής αποδεδειγμένα είναι η πιο ακριβής στην μέτρηση της καταναλισκόμενης ενέργειας είδαμε ότι εμφανίζει ορισμένα μειονεκτήματα όπως είναι το υψηλό κόστος των αισθητήρων για κάθε συσκευή ξεχωριστά και η πολυπλοκότητα της εγκατάστασης. Λαμβάνοντας υπόψη όλα αυτά, ειδικά για την περίπτωση μεγάλων

εφαρμογών ευνοείται η χρήση μη παρεμβατικών μεθόδων όπου και αυτές θα αναλύσουμε στην συνέχεια αυτής της εργασίας. Κατά συνέπεια, η επιστημονική κοινότητα έστρεψε την προσοχή της στην ανάπτυξη και την βελτίωση των μη παρεμβατικών μεθόδων, προκειμένου να καταστεί μια βιώσιμη λύση στην παρακολούθηση του ηλεκτρικού φορτίου.

Η μη-παρεμβατική παρακολούθηση ηλεκτρικών φορτίων (non-intrusive load monitoring) είναι μία τεχνική κατά την οποία διαχωρίζεται η συνολική κατανάλωση ενέργειας ενός κτηρίου σε επιμέρους συσκευές. Η μόνη απαίτηση ενός τέτοιου συστήματος είναι η εγκατάσταση ενός και μόνο αισθητήρα στον κεντρικό ηλεκτρολογικό πίνακα του δικτύου (Εικόνα 1.5). Με αυτό τον τρόπο αναιρείται η ανάγκη της τοποθέτησης πολλών μετρητών μεταξύ των συσκευών και των σημείων τροφοδοσίας τους. Η παραπάνω τεχνική βασίζεται στις μεθόδους επιμερισμού ενέργειας, μια έννοια που θα αναλύσουμε στο επόμενο κεφάλαιο της εργασίας [4].



Εικόνα 1.5: Με την χρήση ενός μετρητή οδηγούμαστε στον επιμερισμό της ενέργειας.

ΚΕΦΑΛΑΙΟ 2

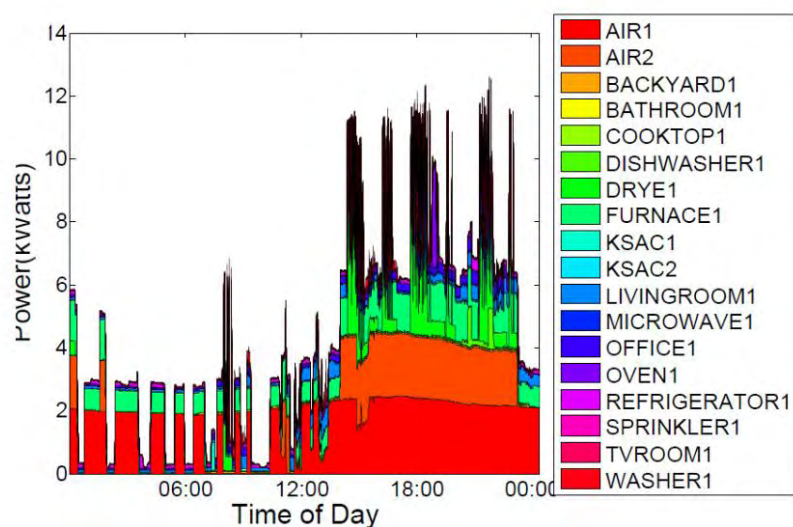
Ο ΕΠΙΜΕΡΙΣΜΟΣ ΤΗΣ ΕΝΕΡΓΕΙΑΣ

2.1 Ορισμός

Ο ενεργειακός επιμερισμός σχετίζεται με την επιστήμη της επεξεργασίας σημάτων και της μηχανικής μάθησης. Με τον όρο επιμερισμό της ενέργειας εννοούμε ότι από το ολικό σήμα της καταναλισκόμενης ενέργειας το οποίο είναι η συνολική ισχύς που καταναλώνουν όλες οι συνδεδεμένες συσκευές σε ένα δίκτυο κατά την διάρκεια της λειτουργίας τους, είναι δυνατό μέσω διάφορων αλγορίθμων να οδηγηθούμε στην εξαγωγή του ενεργειακού προφίλ κατανάλωσης κάθε επιμέρους συσκευής (Εικόνα 2.1).

Πρωταρχικός στόχος των ερευνών που πραγματοποιούνται πάνω στον επιμερισμό της ενέργειας είναι να παρέχει όσο το δυνατόν ακριβέστερες πληροφορίες και δεδομένα, την γνώση της καταναλωτικής συμπεριφοράς κάθε συσκευής τους, γνώση που είναι πολύτιμη τόσο στα χέρια των καταναλωτών όσο και των προμηθευτών.

Στις έρευνες και στην επιστημονική βιβλιογραφία χρησιμοποιούνται αρκετοί όροι που όμως αναφέρονται όλοι τους στην έννοια του επιμερισμού της ενέργειας. Μερικοί όροι από αυτούς δίνονται παρακάτω μαζί με την αγγλική τους ορολογία: Μη παρεμβατική παρακολούθηση φορτίου (Non-intrusive load monitoring), παρακολούθηση φόρτισης μη παρεμβατικών συσκευών (non-intrusive appliance load monitoring), παρακολούθηση φορτίου συσκευής (appliance load monitoring), μη παρεμβατικό σύστημα παρακολούθησης φορτίου συσκευής (non-intrusive appliance load monitoring system) και εμπορικό μη παρεμβατικό σύστημα παρακολούθησης (commercial non-intrusive load monitoring system). Επικρατέστερο και πιο διαδεδομένο όνομα φαίνεται να είναι το Non-Intrusive Load Monitoring (NILM) [5].



Εικόνα 2.1: Χαρακτηριστικό παράδειγμα επιμερισμού της ενέργειας. Από το σήμα της συνολικής κατανάλωσης γίνεται αποσύνθεση της κατανάλωσης κάθε συσκευής ξεχωριστά.

2.2 Η Μεθοδολογία της Μη Παρεμβατικής Παρακολούθησης Φορτίου

Πρωτεργάτης της Μη Παρεμβατικής Παρακολούθησης Φορτίου μπορεί να θεωρηθεί ο Αμερικανός καθηγητής μαθηματικών George W. Hart ο οποίος πριν από περίπου 25 χρόνια είχε την ιδέα αυτής της μεθόδου παρακολούθησης και επισήμως ανέφερε και πρότεινε τον πρώτο αλγόριθμο επιμερισμού πάνω στον οποίο στηρίχτηκαν όλες οι μελλοντικές έρευνες. Το πρόβλημα μπορεί να μοντελοποιηθεί από την παρακάτω μαθηματική εξίσωση:

$$P(t)=p_1(t)+p_2(t)+\dots+p_n(t)$$

όπου η ποσότητα $P(t)$ συμβολίζει την συνολική καταναλισκόμενη ενέργεια των συνδεδεμένων n συσκευών και η $p_i(t)$ την κατανάλωση της i -οστής συσκευής τη χρονική στιγμή t . Το $P(t)$ είναι η μόνη παράμετρος που χρειάζεται να μετρηθεί και αυτό επιτυγχάνεται όπως έχουμε ήδη αναφέρει με έναν μετρητή τοποθετημένο στο σημείο του γενικού ηλεκτρικού πίνακα του ηλεκτρικού δικτύου του χρήστη. Το ζητούμενο λοιπόν είναι η βέλτιστη προσέγγιση των πραγματικών τιμών των μεταβλητών $p_i(t)$, η

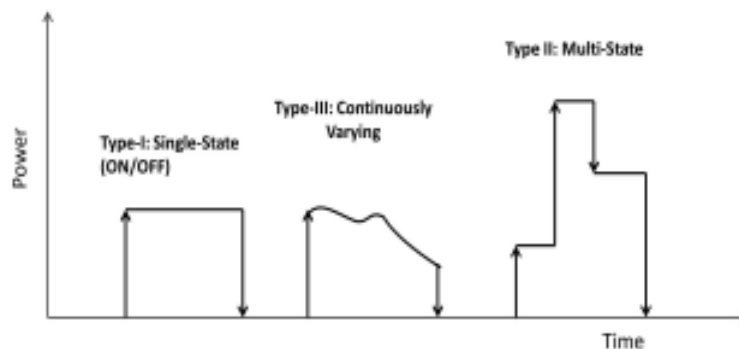
αποσύνθεση δηλαδή της συνολικής κατανάλωσης $P(t)$ στις επιμέρους τιμές. Η λύση αυτού του προβλήματος στηρίζεται στα μοναδικά χαρακτηριστικά της κάθε συσκευής. Όπως κάθε άνθρωπος έχει μοναδικά χαρακτηριστικά, όπως είναι το δαχτυλικό του αποτύπωμα έτσι και κάθε ηλεκτρική συσκευή, το φορτίο καλύτερα κάθε η ηλεκτρικής συσκευής παρουσιάζει ένα συγκεκριμένο προφίλ κατανάλωσης που συναντάμε συχνά να αναφέρεται ως «ηλεκτρικό αποτύπωμα» ή ως «υπογραφή» της συσκευής. Εκμεταλλευόμενοι λοιπόν αυτές τις μοναδικές υπογραφές, το πρόβλημα είναι επιλύσιμο και είναι δυνατή η ανάπτυξη κατάλληλων αλγορίθμων ικανών στην αναγνώριση των επιμέρους καταναλώσεων μέσα στο αθροιστικό σήμα.

Η παρακολούθηση του ηλεκτρικού φορτίου με μη παρεμβατική μέθοδο προϋποθέτει την συλλογή των απαραίτητων δεδομένων, την εξαγωγή των μοναδικών χαρακτηριστικών των συσκευών, δηλαδή τις ηλεκτρικές υπογραφές τους, την μάθηση του αλγορίθμου επιμερισμού με αυτά τα δεδομένα και τελικά την αποσύνθεση της ενέργειας σε επιμέρους τιμές που είναι και ο στόχος της διαδικασίας. Ωστόσο, η διαδικασία παρουσιάζει διαφορές από εφαρμογή σε εφαρμογή, καθώς το περιβάλλον του προβλήματος θέτει περιορισμούς και οι στόχοι μπορεί να είναι διαφορετικοί. Επομένως, αρχικό μέλημα πρέπει να είναι να ληφθούν υπόψη οι παράγοντες που επηρεάζουν το σύστημα. Ερωτήματα που πρέπει να απαντηθούν είναι τα είδη των συσκευών που πρόκειται να τεθούν υπό παρακολούθηση, το είδος του μετρητή με τον οποίο θα γίνεται η μέτρηση της συνολικής ενέργειας και τα δεδομένα που θα παρέχει και τα χαρακτηριστικά γνωρίσματα του σήματος που θα εξάγονται προς χρήση. Οι παράγοντες που αναφέρθηκαν αναλύονται στην συνέχεια [6].

2.2.1 Διάκριση Συσκευών

Όπως έχουμε αναφέρει, η μοναδική ηλεκτρική υπογραφή κάθε ηλεκτρικής συσκευής καθιστά το πρόβλημα της ενεργειακής αποσύνθεσης επιλύσιμο. Οι συσκευές διακρίνονται σύμφωνα με το προφίλ της ενεργειακής τους κατανάλωσης, ωστόσο μπορούν να χρησιμοποιηθούν ταυτόχρονα και άλλα είδη κατηγοριοποίησης. Διακρίνονται τριών ειδών κατηγορίες ως εξής:

1. Τρόπος λειτουργίας: Η λειτουργία ενός φορτίου μπορεί να μην αρκείται σε καταστάσεις ON/OFF αλλά να επιδέχεται και άλλους τρόπους χειρισμού. Έτσι οι συσκευές με κριτήριο τον τρόπο λειτουργίας τους διακρίνονται στους παρακάτω τύπους (Εικόνα 2.2):
 - I. Type-1: Συσκευές αυτής της κατηγορίας παρουσιάζουν μόνο δυο καταστάσεις λειτουργίας που αναφέρονται ως ON/OFF. Με άλλα λόγια, μια συσκευή type-1 είτε μπορεί να είναι σε λειτουργία είτε εκτός λειτουργίας, χωρίς να υπάρχει κάποια ενδιάμεση κατάσταση. Παράδειγμα συσκευής τέτοιου τύπου μπορεί να είναι ένας λαμπτήρας ο οποίος είτε ανάβει είτε όχι.
 - II. Type-2: Σε αυτήν την κατηγορία συναντάμε συσκευές που δεν αρκούνται σε δύο καταστάσεις λειτουργίας αλλά παρουσιάζουν επιπλέον ένα πεπερασμένο πλήθος καταστάσεων που εμφανίζουν μοναδικά μοτίβα. Τέτοια συσκευή είναι η ηλεκτρική κουζίνα η οποία ανάλογα με την ρύθμιση του θερμοστάτη της αλλάζει αντίστοιχα την λειτουργία της.
 - III. Type-3: Η μόνη διαφορά των συσκευών που περιλαμβάνονται σε αυτήν την κατηγορία από αυτές των type-2 είναι ότι διαθέτουν άγνωστο ή μη προσεγγίσιμο αριθμό καταστάσεων λειτουργίας. Χαρακτηριστικό παράδειγμα ο ηλεκτρονικός προσωπικός υπολογιστής εξαιτίας της φύσης και της πολυπλοκότητας των συστημάτων του παρουσιάζει απεριόριστες καταστάσεις λειτουργίας που είναι αδύνατον να καταγραφούν και να ανιχνευθούν.
 - IV. Type-4: Εδώ συναντάμε συσκευές που παραμένουν ενεργές για πολύ μεγάλα χρονικά διαστήματα, μέρες ή ακόμα και εβδομάδες και καταναλώνουν ενέργεια με σταθερό ρυθμό. Χάρης σε αυτό το χαρακτηριστικό τους αναφέρονται και ως «μονίμως καταναλωτικές συσκευές». Το ρούτερ ανήκει σε αυτήν την κατηγορία μιας και είναι μονίμως σε λειτουργία και δεν έχει μεταβαλλόμενη ισχύ [7].

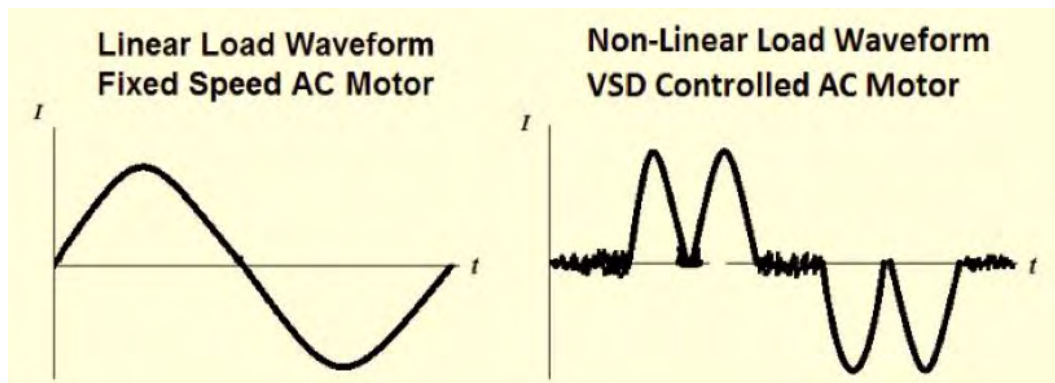


Εικόνα 2.2: Είδη φορτίων βασισμένα στο μοτίβο κατανάλωσης ενέργειας που παρουσιάζουν.

2. Κλίμακα κατανάλωσης ισχύος: Η ηλεκτρική ενέργεια και ο ρυθμός με τον οποία καταναλώνεται διαφοροποιείται από συσκευή σε συσκευή και έτσι μπορούμε να διακρίνουμε συσκευές που καταναλώνουν μερικά Watt, όπως οι λάμπες τεχνολογίας LED και συσκευές μερικών kWatt δηλαδή μεγάλης κατανάλωσης ισχύος όπως οι θερμοσήφωνες ή οι ηλεκτρικές σόμπες.
3. Ηλεκτρικά χαρακτηριστικά: Η τελευταία και αυτή που καθορίζει την ηλεκτρική υπογραφή ενός φορτίου κατηγορία είναι τα ηλεκτρικά χαρακτηριστικά ή όπως λέμε η φύση του φορτίου (Εικόνα 2.3). Εδώ συναντάμε τέσσερις υποκατηγορίες:
 - I. Γραμμικά φορτία: Σε αυτήν την κατηγορία συναντάμε συσκευές που περιέχουν συνδιασμούς γραμμικών παθητικών στοιχείων κυκλωμάτων και είναι ανεξάρτητα από τον χρόνο. Η αναπαράσταση τέτοιων φορτίων μπορεί να γίνει με μία απλή ωμική αντίσταση. Δεν παρουσιάζουν μεταβατική συμπεριφορά και στην μόνιμη κατάσταση το ρεύμα εισόδου είναι ημιτονοειδή.
 - II. Ηλεκτρομαγνητικά φορτία: Οι συσκευές αυτές περιλαμβάνουν λείτουργίες που εξαρτώνται από ηλεκτρικούς κινητήρες, όπως μονοφασικούς επαγωγικούς κινητήρες και κινητήρες Universal. Τέτοιες συσκευές είναι οι ηλεκτρικές σκούπες, ανεμιστήρες κτλ. Εδώ το μεταβατικό φαινόμενο με το

που ξεκινάει την λειτουργία της η συσκευή είναι έντονο και χρειάζεται χρόνος για την απόσβεσή του.

- III. Φορτία ηλεκτρονικής φύσης: Σε φορτία αυτού του είδους, το ρεύμα έχει πλούσιο αρμονικό περιεχόμενο και την ύπαρξη ανώτερων αρμονικών συνιστωσών. Φυσικά σε αυτήν την κατηγορία ανήκουν συσκευές που διαθέτουν ηλεκτρονικά στοιχεία, όπως οι υπολογιστές [3].



Εικόνα 2.3: Διαφορετικοί τύποι φορτίων και η αναπαράσταση του ρεύματος στην είσοδο τους.

2.2.2 Μονάδες Απόκτησης Δεδομένων

Στη παρακολούθηση φορτίου με μη παρεμβατική τεχνική η μόνη συσκευή καταγραφής είναι ένας μετρητής ο οποίος τοποθετείται ενδιάμεσα του δικτύου που θέλουμε να μετρήσουμε και του γενικού του πίνακα. Βασική αρμοδιότητα είναι η συλλογή όλων αυτών των απαραίτητων δεδομένων με επαρκή ρυθμό που θα χρησιμοποιηθούν μετέπειτα στους αλγόριθμους αποσύνθεσης της ενέργειας, ώστε να αναγνωριστούν διακριτά μοτίβα φορτίου. Αποτελεί λοιπόν, ένα πολύ σημαντικό κομμάτι του συστήματος.

Τα δεδομένα είναι διακριτά και η δειγματοληψία με την οποία τα αποκτούμε πρέπει να είναι μεγαλύτερη από ένα συγκεκριμένο όριο το οποίο καθορίζεται από την

συχνότητα του ρεύματος του δικτύου. Σύμφωνα με το θεώρημα Nyquist-Shannon, η ελάχιστη επιτρεπόμενη συχνότητα δειγματοληψίας ενός σήματος προς ψηφιοποίηση, η οποία δεν επιφέρει αλλοίωση της πληροφορίας πρέπει να είναι διπλάσια από την μεγαλύτερη συχνότητα των συνιστωσών του σήματος. Λαμβάνοντας υπόψη ότι στο ελληνικό δίκτυο ηλεκτρικού ρεύματος, συναντάμε συχνότητα των 50Hz και ότι για την καταγραφή την καταγραφή των διάφορων παροδικών αλλαγών στην κατανάλωση, ο μετρητής της συνολικής κατανάλωσης θεωρητικά πρέπει να είναι σχεδιασμένος για δειγματοληψία σε συχνότητες της τάξεως των 2kHz. Ωστόσο με μεγαλύτερες συχνότητες συνεπάγεται και περισσότερη πληροφορία αλλά με αύξηση του κόστους του μετρητή αντίστοιχα [8].

Υπάρχει μια μεγάλη ποικιλία μετρητών ενέργειας σχεδιασμένων να μετρούν την καταναλισκόμενη ενέργεια ενός δικτύου οι οποίοι μπορούν να κατηγοριοποιηθούν σε μετρητές υψηλής δειγματοληψίας και μετρητές χαμηλής δειγματοληψίας. Ο ρυθμός δειγματοληψίας καθορίζει τον τύπο των πληροφοριών που μπορούν να εξαχθούν από τα ηλεκτρικά σήματα. Προκειμένου να συλληφθούν οι αρμονικές ανώτερης τάξης των ηλεκτρονικών σημάτων, τα οποία είναι ακέραια πολλαπλάσια της βασικής συχνότητας (50 Hz), ο ρυθμός δειγματοληψίας πρέπει να πληροί τα κριτήρια Nyquist-Shannon. Για παράδειγμα, ένας μετρητής ενέργειας όπως ο Intron έχει ρυθμό δειγματοληψίας 600 Hz και αυτό του επιτρέπει να συλλάβει μέχρι και την 5^η αρμονική ταλάντωση των ηλεκτρικών σημάτων. Από την άλλη πλευρά, οι παραδοσιακές μετρήσεις ισχύος, όπως η πραγματική ισχύς, η ενεργός ισχύς, η τάση της μέσης τετραγωνικής τάσης και οι τιμές ρεύματος μπορούν να υπολογιστούν με χαμηλό ρυθμό δειγματοληψίας της τάξεως των 120 Hz. Οι υπολογισμένες μετρήσεις είτε στέλνονται στο σύστημα υποστήριξης μέσω μιας κάρτας δικτύου, είτε υποβάλλονται σε επεξεργασία μέσα στον μετρητή. Οι κάρτες δικτύου μπορούν να διαβάζουν, να γράφουν και να στέλνουν δεδομένα έως και 1 KHz, ωστόσο απαιτούνται τροποποιήσεις στο υλικό του μετρητή. Για να καταγραφούν τα παροδικά γεγονότα ή ο ηλεκτρικός θόρυβος που παράγεται από τα ηλεκτρικά σήματα, οι κυματομορφές πρέπει να λαμβάνουν δειγματοληψία με πολύ υψηλότερη συχνότητα σε κλίμακα 10 έως 100 MHz. Αυτοί οι τύποι μετρητών ενέργειας υψηλής συχνότητας είναι συχνά κατασκευασμένοι κατά παραγγελία και είναι προσαρμοσμένοι στο είδος των χαρακτηριστικών που χρειάζεται να εξαχθούν.

Οι ερευνητές υποστηρίζουν ότι οι περισσότεροι εμπορικά διαθέσιμοι μετρητές παρουσιάζουν διαφορές στις μετρήσεις των δεδομένων τους από 10% έως 20%. Ακόμα, οι χαμηλού κόστους λύσεις μέτρησης προσφέρουν περιορισμένες λειτουργίες, όπως μετατροπές αναλογικού σήματος σε ψηφιακό χαμηλής ανάλυσης και κάρτα μνήμης μικρού μεγέθους που χρησιμοποιείται από την μονάδα επεξεργασίας για την αποθήκευση αποτελεσμάτων. Έτσι λοιπόν, για να επιτύχουμε δειγματοληψία μεσαίων ή υψηλότερων επιπέδων των ηλεκτρικών σημάτων, οι ερευνητές άρχισαν να αναπτύσσουν τα δικά τους πρωτότυπα για πειραματικές αξιολογήσεις. Η απόκτηση δεδομένων για το NILM μπορεί περαιτέρω να κατηγοριοποιηθεί σε δεδομένα ολόκληρου του σπιτιού και του κυκλώματος. Το τυπικό σύστημα NILM χρησιμοποιεί δεδομένα ολόκληρης οικίας που αποκτήθηκαν από έναν μόνο μετρητή. Όμως, ένας περιορισμός μιας τέτοιας προσέγγισης είναι ότι ο εντοπισμός συσκευών χαμηλής ισχύος και μεταβλητών συσκευών με την παρουσία φορτίων μεγάλης ισχύος από τα δεδομένα ολόκληρης της οικίας, που συχνά γίνεται αρκετά δύσκολος. Για αυτό έχει προταθεί μια εναλλακτική προσέγγιση, σύμφωνα με την οποία χρησιμοποιούνται μετρήσεις ισχύος σε επίπεδο κυκλώματος, με τις συσκευές υψηλής κατανάλωσης να λαμβάνουν ένα ειδικό κύκλωμα εντός των σπιτιών. Συνεπώς το έργο της αποσύνθεσης ισχύος γίνεται πολύ πιο εύκολο, καθώς υπάρχουν λιγότερες συσκευές σε κάθε κύκλωμα σε αντίθεση με ολόκληρο το σπίτι, αλλά με αποτέλεσμα το αυξημένο κόστος εγκατάστασης και πολυπλοκότητας [6].

2.2.3 Εξαγωγή Χαρακτηριστικών

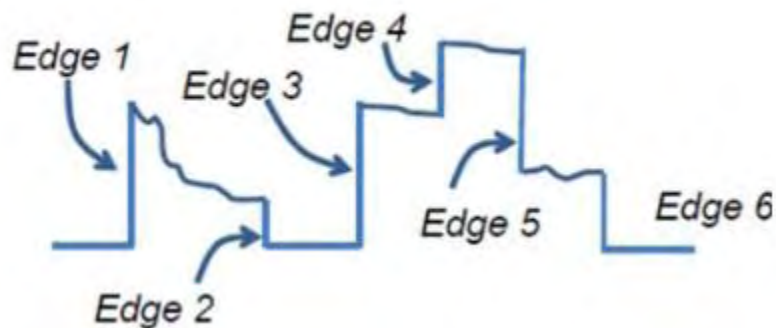
Το βήμα που ακολουθεί κατόπιν της συλλογής δεδομένων είναι η επεξεργασία αυτών. Τα δεδομένα που λαμβάνονται έχουν να κάνουν με τάσεις και κυματομορφές ρεύματος και χρίζουν επεξεργασία ώστε να έρθουν σε μορφή κατάλληλη προκειμένου να υπολογιστούν οι μετρήσεις ισχύος δηλαδή να εξαχθούν οι μετρήσεις ενεργούς και αέργου ισχύος. Στην συνέχεια έρχεται η ανίχνευση συμβάντων (events) που μπορεί να είναι η σύνδεση ή αποσύνδεση μιας συσκευής στο δίκτυο ή η μεταγωγή της κατάσταση μιας άλλης. Μια μονάδα ανίχνευσης συμβάντος έχει ως ρόλο την ανίχνευση της μετάβασης των συσκευών του δικτύου που επιτηρεί σε λειτουργία ON/OFF με ανάλυση

των αλλαγών στα επίπεδα ισχύος. Μια τέτοια μονάδα στοχεύει στον εντοπισμό μεταβολών στην κατανάλωση με σκοπό την ανίχνευση ενεργοποίησης ή απενεργοποίησης συσκευής τύπου I. Για τις συσκευές των υπόλοιπων τύπων μια τέτοια τακτική δεν θα ήταν δυνατή για αυτό και σε αυτές τις περιπτώσεις σειρά παίρνει η εκμετάλλευση των ηλεκτρικών υπογραφών τους. Οι προσεγγίσεις βασίζονται σε γεγονότα στηρίζονται στο ότι η καταναλισκόμενη ισχύς ενός σπιτιού αλλάζει διαρκώς εμφανίζοντας αυξήσεις και πτώσεις ανάλογα των συσκευών που λειτουργούν (Εικόνα 2.4) [9]. Η διαδικασία ανίχνευσης γεγονότων σε ένα ηλεκτρικό δίκτυο μπορεί να επιτευχθεί με δύο κύριες μεθόδους: την ανάλυση σταθερών καταστάσεων και την ανάλυση μεταβλητών καταστάσεων.

Στην μεθοδολογία ανάλυσης σταθερών καταστάσεων μέλημα είναι ο εντοπισμός ενεργοποίησης και απενεργοποίησης συσκευών. Σε αυτήν την περίπτωση ηλεκτρικά μεγέθη όπως η πραγματική και άεργη ισχύς, η ένταση και η τάση ρεύματος φαίνονται χρήσιμα. Η πραγματική ισχύς και η αεργός ισχύς είναι οι δυο από τα πιο διαδεδομένα μεγέθη για τον προσδιορισμό της κατάστασης μιας συσκευής. Η πραγματική ισχύς είναι ο ρυθμός με τον οποίο καταναλώνει ενέργεια μια συσκευή κατά την διάρκεια της λειτουργίας της. Με την χρήση της πραγματικής ισχύς εντοπίζονται εύκολα συσκευές τύπου I και ορισμένες τύπου II που παρουσιάζουν υψηλή κατανάλωση. Εδώ απαιτείται μικρή συχνότητα δειγματοληψίας επομένως χαμηλότερο κόστος για την μέτρηση και επεξεργασία του σήματος. Μια πειραματική και ενναλακτική προσέγγιση είναι η χρήση τιμών έντασης και τάσης των συσκευών και η εξαγωγή χαρακτηριστικών όπως διαφορά φάσης και λόγος ισχύος. Αν το η συσκευή περιέχει καθαρά αντίσταση οι κυματομορφές ρεύματος και τάσης δεν θα παρουσιάζουν διαφορά φάσης. Αν περιέχει επαγωγικά και χωρητικά στοιχεία, εμφανίζεται διαφορά φάσης και παράγεται ή καταναλώνεται άεργη ισχύς. Με αυτό τον τρόπο μπορούν να εντοπιστούν και να διακριθούν φορτία με αντιστατικές συμπεριφορές και φορτία με επαγωγικές ή χωρητικές συμπεριφορές. Βέβαια σε αυτήν την περίπτωση απαιτείται μεγαλύτερη συχνότητα δειγματοληψίας άρα και υψηλότερο κόστος και εξαιρώντας τις συσκευές τύπου III.

Στην μεθοδολογία ανάλυσης μεταβλητών καταστάσεων γίνεται χρήση περισσότερων χαρακτηριστικών μιας και οι συσκευές μεταβλητής κατάστασης το επιτρέπουν. Με λίγα λόγια χρησιμοποιούν τις μεταβατικές υπογραφές που διακρίνουν

μοναδικά τις συσκευές με τις χαρακτηριστικές μεταβάσεις της κατάστασής τους. Σε μία τέτοια μεθοδολογία είναι συνηθής ο μετασχηματισμός Fourier για την χρήση του συχνοτικού περιεχομένου των σημάτων και των διαχωρίσιμων χαρακτηριστικών που θα επιτρέψουν την ευκολότερη παρακολούθησή τους. Οι συσκευές μεταβλητών καταστάσεων έχουν το ιδιαίτερο χαρακτηριστικό να επιστρέφουν θόρυβο στην γραμμή τροφοδοσία τους κατά την ενεργοποίηση ή την απενεργοποίησή τους. Μπορεί λοιπόν να χρησιμοποιήσουμε τις υπογραφές θορύβου που επιπλέον μένουν σταθερές κατά την διάρκεια της λειτουργίας των συσκευών ώστε να ανιχνεύονται τα γεγονότα των συσκευών μεταβλητών καταστάσεων. Φυσικά το κόστος αυξάνεται λόγω της υψηλής συχνότητας δειγματοληψίας [8].

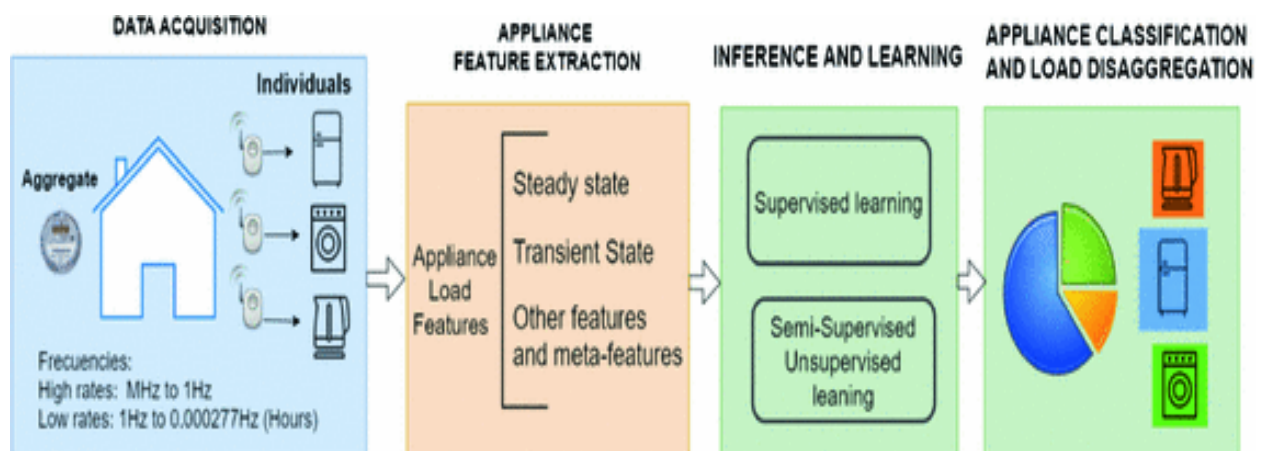


Εικόνα 2.4: Σχηματική περιγραφή της ανίχνευσης συμβάντων.

2.2.4 Αναγνώριση Φορτίου

Τα εξαγόμενα χαρακτηριστικά των συσκευών στην συνέχεια αναλύονται περαιτέρω από αλγόριθμους αναγνώρισης φορτίου. Σε αυτό λοιπόν το στάδιο χρησιμοποιούνται μαθηματικά μοντέλα για την αποσύνθεση του σήματος της συνολικής καταναλισκόμενης ενέργειας σε επιμέρους καταναλώσεις. Οι κυριότεροι αλγόριθμοι που έχουν αξιοποιηθεί είναι αλγόριθμοι βελτιστοποίησης και αναγνώρισης προτύπων με την επιστημονική κοινότητα να στηρίζεται κυρίως στην αναγνώριση προτύπων εξαιτίας της πολυπλοκότητας που παρουσιάζουν οι μέθοδοι βελτιστοποίησης. Η ανάγκη μεγάλου όγκου δεδομένων για την εκπαίδευση των αλγορίθμων είναι ένα σημαντικό εμπόδιο που οι ερευνητές καλούνται να αντιμετωπίσουν. Η εκπαίδευση των αλγορίθμων είτε γίνεται

με επιτηρούμενη είτε με μη-επιτηρούμενη μάθηση, δύο έννοιες τις οποίες θα αναλύσουμε στον επόμενο κεφάλαιο. Με την εξέλιξη της επιστήμης, οι ερευνητές έχουν στα χέρια τους ποικίλα εργαλεία και μπορούν να προσεγγίσουν το ζήτημα της αναγνώρισης φορτίου με διάφορες μεθόδους προερχόμενες από διάφορους τομείς της επιστήμης, με κάθε μια να έχει τα πλεονεκτήματα και τα μειονεκτήματά της. [9]. Με τα παραπάνω βήματα καταλήγουμε τελικά στην αποσύνθεση του φορτίου (Εικόνα 2.5).



Εικόνα 2.5: Τα βήματα της μεθόδου NILM.

ΚΕΦΑΛΑΙΟ 3

ΜΕΘΟΔΟΙ NILM

3.1 Βασικές Έννοιες

Η παρακολούθηση φορτίου με μη-παρεμβατική τεχνική είναι η επικρατέστερη μέθοδος που χρησιμοποιείται για την επιτήρηση του ενεργειακού προφίλ μιας κατοικίας και τον κατακερματισμό της συνολικής της κατανάλωσης ενέργειας στις επιμέρους συσκευές της. Ενώ οι πιο δημοφιλείς αλγόριθμοι αποσύνθεσης ενέργειας βασίζονται στα κρυπτομαρκοβιανά μοντέλα (Hidden Markov Model), λύσεις προερχόμενες από τον χώρο των νευρωνικών δικτύων έχουν προσελκύσει το ενδιαφέρον της ερευνητικής κοινότητας [10]. Πριν περάσουμε όμως στην περαιτέρω αναφορά και ανάλυση μεθόδων, στο σημείο αυτό θα ήταν χρήσιμο να επισημάνουμε κάποιες βασικές έννοιες.

3.1.1 Μηχανική Μάθηση

Η μηχανική μάθηση είναι ένα υποπεδίο της επιστήμης των υπολογιστών που μελετά τον τρόπο με τον οποίο οι υπολογιστικές μηχανές θα μπορούν να μιμούνται την ικανότητα των ανθρώπων στην μάθηση ή πιο απλά να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί [11]. Αντικείμενο μελέτης της μηχανικής μάθησης είναι η διερεύνηση και η κατασκευή αλγορίθμων ικανών να μαθαίνουν από δεδομένα και να προβαίνουν σε προβλέψεις ή να εξάγουν αποφάσεις χαρακτηριζόμενες ως αποτελέσματα. Ωστόσο, η μηχανική μάθηση ανάλογα με τον τρόπο εκπαίδευσης του αλγορίθμου και της μεθόδου ανατροφοδότησής του διακρίνεται σε 3 κύριες κατηγορίες [12]:

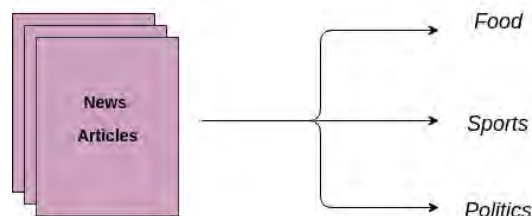
- Επιτηρούμενη μάθηση ή επιβλεπόμενη μάθηση (supervised learning). Οι αλγόριθμοι αυτής της μεθόδου προϋποθέτουν την εκπαίδευσή τους με γνωστά παραδείγματα δεδομένων εισόδου-εξόδου. Για να επιτευχθεί μια επιτυχημένη προσέγγιση της συνάρτησης στόχος, δηλαδή ενός γενικού κανόνα που θα αντιστοιχεί τις εισόδους με τις εξόδους απαιτείται η τροφοδοσία του αλγορίθμου

με ένα σύνολο εισόδων που καλούνται στιγμιότυπα, που περιγράφονται από κάποια χαρακτηριστικά και έχουν γνωστές τιμές-εξόδους.

- Μη επιτηρούμενη μάθηση ή μη επιβλεπόμενη μάθηση (supervised learning). Οι αλγόριθμοι αυτού του είδους διαφέρουν από το προηγούμενο στο γεγονός ότι δεν χρειάζονται την διάθεση ενός συνόλου δεδομένων εκπαίδευσης με γνωστά τα ζεύγη εισόδων-εξόδων. Εδώ, θα πρέπει να ανακαλυφθούν οι συσχετίσεις των ομάδων δεδομένων, από τα χαρακτηριστικά τους.
- Ενισχυτική μάθηση (reinforcement learning). Ο αλγόριθμος αλληλεπιδρά με το περιβάλλον του το οποίο είναι δυναμικό και επιτυγχάνει το στόχο του, χωρίς ωστόσο να του έχουμε θέσει ρητά κατά πόσο τον προσεγγίζει. Τέτοιοι αλγόριθμοι χρειάζονται αρκετές εκπαιδεύσεις ώστε να γίνουν αποτελεσματικοί.

Παράλληλα, οι αλγόριθμοι μηχανικής μάθησης μπορούν να κατηγοριοποιηθούν με βάση τον στόχο που τους έχει τεθεί.

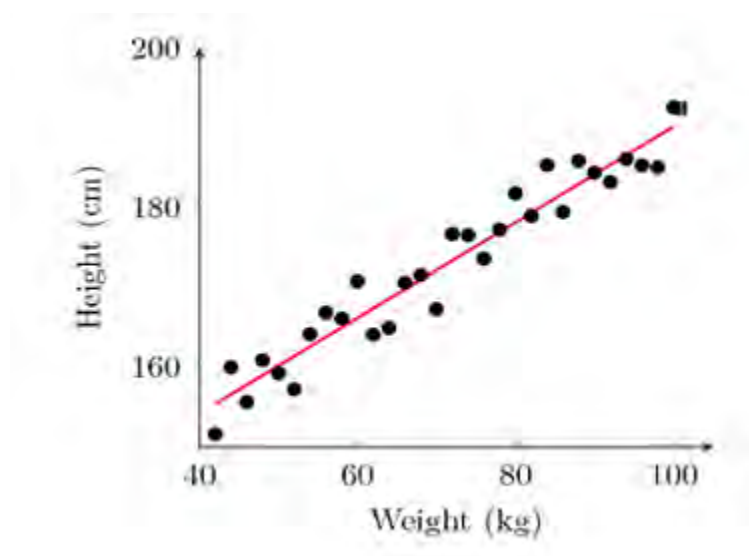
- Ταξινόμηση (Classification) Σε αυτήν την περίπτωση τα δεδομένα διακρίνονται από τουλάχιστον δύο κλάσεις και ο αλγόριθμος θα πρέπει να τα τοποθετήσει σε μία ή περισσότερες κλάσεις(Εικόνα 3.1).



Εικόνα 3.1: Τα δεδομένα- άρθρα που μπορεί να είναι οποιουδήποτε είδους πρέπει να ταξινομηθούν σε τρεις κλάσεις- είδη άρθρων.

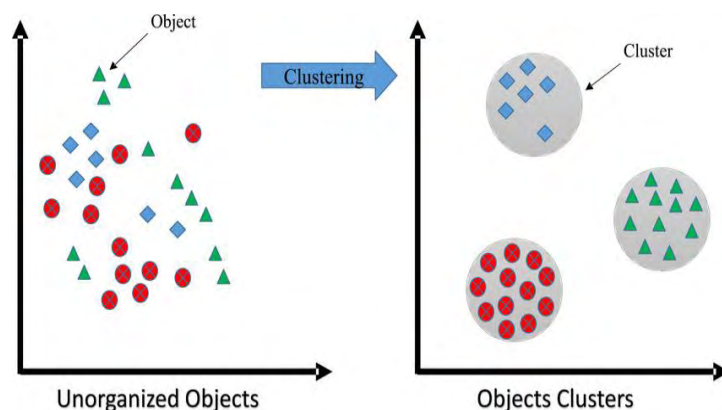
- Παλινδρόμηση (Regression). Ερευνάτε η συσχέτιση μεταξύ μιας εξαρτώμενης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Απαίτηση της τεχνικής αυτής είναι ένα πλήθος γνωστού ζεύγους δεδομένων της μιας εξαρτώμενης μεταβλητής με το αντίστοιχη τιμή της ανεξάρτητης ώστε να

καθοριστεί η καλύτερη συνάρτηση μοντελοποίησης των δεδομένων (Εικόνα 3.2).



Εικόνα 3.2: Με βάση τα δοσμένα δεδομένα δημιουργείται η καλύτερη συνάρτηση (κόκκινη γραμμή) που προσεγγίζει αυτά αλλά ικανή να προβλέψει και μελλοντικά.

- Ομαδοποίηση (Clustering). Σε αυτήν την διαδικασία ένα σύνολο δεδομένων-αντικειμένων διαχωρίζεται σε ένα σύνολο λογικών ομάδων με γνώμονα την ομοιότητά τους (Εικόνα 3.3).

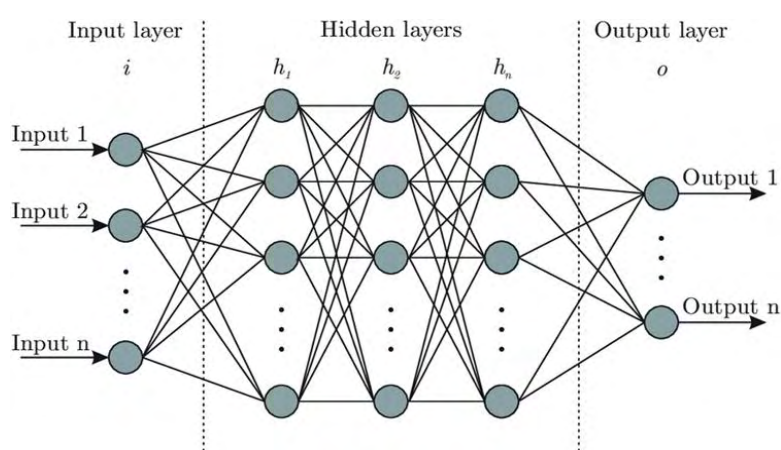


Εικόνα 3.3: Αντικείμενα με όμοια χαρακτηριστικά διαχωρίζονται σε κατάλληλες ομάδες.

3.1.2 Νευρωνικά Δίκτυα

Ο ανθρώπινος εγκέφαλος συγκροτείται από μια τεράστια ποικιλία νευρώνων διασυνδεδεμένων μεταξύ τους. Ένας βιολογικός νευρώνας είναι ένα κύτταρο

προικισμένο με την δυνατότητα μετάδοσης ενός ηλεκτρικού σήματος. Αποτελείται κατά κύριο λόγο από μια διακλαδωτική διάρθρωση εισροών (δενδρίτες) που έχει τον ρόλο της εισόδου των σημάτων, ένα κυτταρικό σώμα και μια διακλαδωτική δομή εκροών (άξονες) για την έξοδο των σημάτων. Οι άξονες ενός νευρώνα είναι συνδεδεμένοι με τους δενδρίτες ενός άλλου. Το κύτταρο ενεργοποιείται και μεταδίδει το ηλεκτροχημικό σήμα στους άξόνες του, όταν το σήμα που λάβει μέσω των δενδριτών ξεπεράσει ένα όριο. Η παραπάνω δομή και λειτουργία έδωσε την έμπνευση στην υπολογιστική επιστήμη για την κατασκευή ενός αφηρημένου συνόλου αλγορίθμων, με στόχο την επίλυση περίπλοκων υπολογιστικών προβλημάτων. Στην περίπτωση λοιπόν του τεχνητού νευρωνικού δικτύου, έχουμε να κάνουμε με μια προσομοίωση της δομής και της λειτουργίας του ανθρώπινου εγκεφάλου, όπου οι νευρώνες είναι κόμβοι που δέχονται σύνολα αριθμητικών εισόδων είτε από άλλους νευρώνες είτε από το περιβάλλον, τα επεξεργάζονται και παράγουν αντίστοιχα σήματα εξόδου. Ανάλογα με τις διασυνδέσεις τους οι νευρώνες διακρίνονται σε νευρώνες εισόδου, εξόδου και σε υπολογιστικούς ή αλλιώς κρυμμένους νευρώνες (Εικόνα 3.4). Συγκεκριμένα, οι νευρώνες εισόδου και εξόδου, συνδέονται με τους υπολογιστικούς νευρώνες και παίζουν τον ρόλο του μεσολαβητή του νευρωνικού δικτύου με το περιβάλλον του.



Εικόνα 3.4: Η δομή ενός τεχνητού νευρωνικού δικτύου νευρώνων.

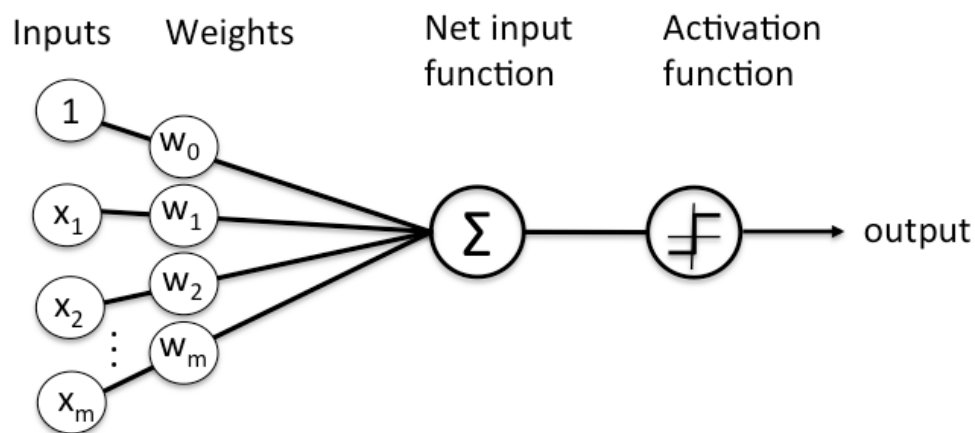
Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδο που δέχονται με μία ξεχωριστή μεταβλητή που ονομάζεται συναπτικό βάρος και είναι μοναδική για κάθε νευρώνα και στην συνέχεια υπολογίζουν το άθροισμα όλων αυτών των γινομένων. Η

ποσότητα αυτή στην συνέχεια γίνεται όρισμα της συνάρτησης ενεργοποίησης, την οποία υπολογίζει κάθε κόμβος και τροφοδοτεί με την τιμή τις εξόδους του.

Η έξοδος y_k του k νευρώνα του δικτύου δίνεται από την εξίσωση:

$$y_k = \varphi \left(\sum_{i=0}^N x_{ki} w_{ki} \right)$$

όπου η x_{ki} είναι η i -οστή είσοδος του k νευρώνα, w_{ki} το i -οστό συναπτικό βάρος του k νευρώνα και $\varphi(\cdot)$ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου. Ιδιαίτερη σημασία παρουσιάζει το w_{k0} , το οποίο ονομάζεται πόλωση (bias), η τιμή της εισόδους ισούται πάντα με την μονάδα και αν το άθροισμα των υπόλοιπων εισόδων του νευρώνα είναι μεγαλύτερο από την τιμή αυτή, τότε πυροδοτείται η ενεργοποίηση του νευρώνα [13] (Εικόνα 3.5).



Εικόνα 3.5: Διάγραμμα που απεικονίζει την λειτουργία ενός νευρώνα.

3.2 Μαρκοβιανά Μοντέλα

3.2.1 Αλυσίδες Markov

Οι αλυσίδες Markov (Markov chains) είναι στοχαστικές διαδικασίες χωρίς την δυνατότητα της μνήμης που έχουν ένα πεπερασμένο αριθμό καταστάσεων. Μια διαδικασία είναι ένα σύστημα του οποίου η κατάσταση παριστάνεται από μια ένα σύνολο δεδομένων. Μια στοχαστική διαδικασία είναι μια οικογένεια τυχαίων μεταβλητών $\{X(t) : t \in T\}$, όπου t είναι μία παράμετρος που παίρνει τιμές σε ένα κατάλληλα ορισμένο σύνολο T . Με την έννοια «χωρίς μνήμη» αναφερόμαστε στο γεγονός ότι η κατάσταση του συστήματος σε οποιαδήποτε χρονική στιγμή t κι αν βρίσκεται εξαρτάται μόνο από την κατάσταση του συστήματος την αμέσως προηγούμενη στιγμή $t-1$. Ακόμα θεωρούμε ότι οι αλυσίδες Markov λειτουργούν σε διακριτές χρονικές στιγμές.

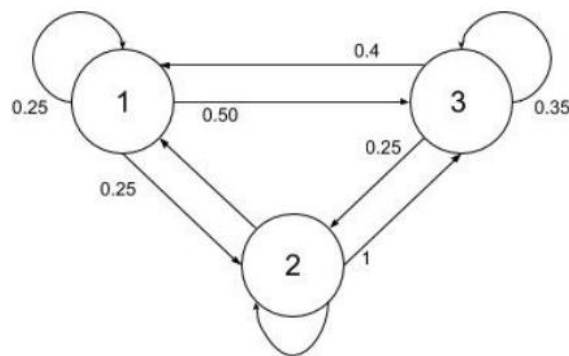
Οι αλυσίδες Markov μπορούν να αναπαρασταθούν ως ένα κατευθυνόμενο γράφημα, στο οποίο κάθε κόμβος αντιστοιχεί σε μια από τις πιθανές καταστάσεις που ανήκουν σε ένα σύνολο καταστάσεων γνωστό ως $S = \{1, 2, \dots, n\}$. Σε κάθε χρονικό βήμα, το μοντέλο έχει μια κατάσταση από το σύνολο S και στην συνέχεια μεταβαίνει σε άλλη, με μια ενέργεια που ονομάζεται transition. Όπως αναφέρθηκε παραπάνω, οι αλυσίδες Markov είναι στοχαστικές διαδικασίες, συνεπώς δεδομένης της κατάστασης του μοντέλου, οι πιθανές μεταβάσεις ακολουθούν μια ορισμένη κατανομή πιθανότητας. Συγκεκριμένα, κάθε μετάβαση P_{ij} αντιστοιχεί σε μια από τις άκρες του γραφήματος με το οποίο αναπαρίστανται και ορίζεται ως η πιθανότητα μιας πιθανής μετάβασης από την κατάσταση i στην κατάσταση j . Μετάβαση μπορεί να θεωρηθεί ότι έχουμε ακόμα και αν παραμείνει στην ίδια κατάσταση. Από την στιγμή που οι μεταβάσεις αντιπροσωπεύουν πιθανότητες τότε για το σύνολό τους θα πρέπει να ισχύει:

$$\sum_j P_{ij} = 1$$

Οι πιθανότητες των μεταβάσεων συλλέγονται στον πίνακα μεταβάσεων P , ο οποίος είναι ένας πίνακας μεγέθους $n \times n$ και αποθηκεύει την τιμή κάθε πιθανότητας P_{ij} στην αντίστοιχη θέση του. Ακόμα, οι αλυσίδες Markov περιλαμβάνουν μια αρχική κατανομή της κατάστασης, που δείχνει την κατάσταση του συστήματος την χρονική στιγμή $t=1$, και υποδεικνύεται με το διάνυσμα πιθανοτήτων $p(X_1)$. Έτσι, με αυτόν τον τρόπο οι αλυσίδες Markov μπορούν να αναπαρασταθούν πλήρως από τις ποσότητες $\{S, P, p(X_1)\}$. Για παράδειγμα, δοσμένης της αρχικής κατάστασης X_1 και του πίνακα μεταβάσεων P , είναι δυνατό να υπολογιστεί μια κατανομή καταστάσεων την αμέσως επόμενη χρονική στιγμή $t=2$ από το γινόμενο X_1P , αλλά και καταστάσεις οποιασδήποτε χρονικής στιγμής $t=k$ από το γινόμενο X_1P^k . Σύμφωνα με τα παραπάνω, οδηγούμαστε στον εξής τύπο:

$$P(X_{t+1} = x \mid X_t = x_t) = P(X_{t+1} = x \mid X_0 = x_0, X_1 = x_1, \dots, X_t = x_t)$$

Ένα χαρακτηριστικό παράδειγμα αλυσίδων Markov με σύνολο καταστάσεων $S = \{1, 2, 3\}$ και τις πιθανότητες κάθε μετάβασης αναγραφμένες φαίνεται στο παρακάτω γράφημα (Εικόνα 3.6):



Εικόνα 3.6: Αλυσίδες Markov αναπαριστάμενες από γράφημα.

Στο παραπάνω παράδειγμα ο πίνακας μεταβάσεων συμπληρώνεται ακολούθως:

$$P = \begin{pmatrix} 0.25 & 0.25 & 0.5 \\ 0 & 0 & 1 \\ 0.4 & 0.25 & 0.35 \end{pmatrix}$$

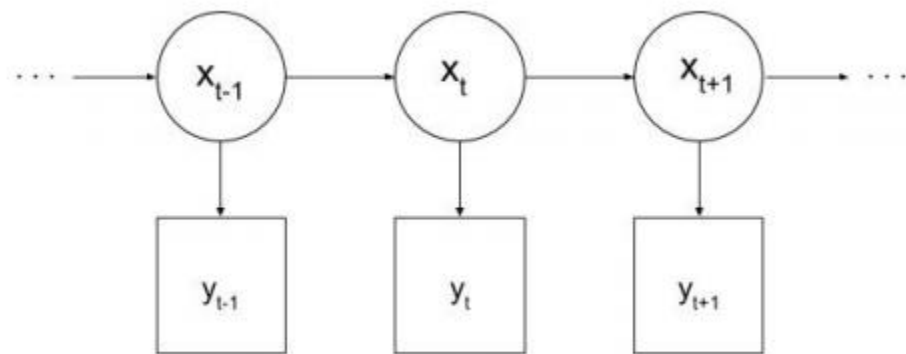
Έτσι λοιπόν, για μια αρχική κατάσταση $\chi_1 = (0.5, 0.5, 0)$ στην επόμενη χρονική στιγμή $t = 1$ υπολογίζουμε $\chi_1 P = (0.125, 0.125, 0.75)$ και συνεχίζουμε αναλόγως για κάθε χρονική στιγμή t .

Οι αλυσίδες Markov είναι μια ειδική περίπτωση των διαδικασιών Markov, έτσι ονομάζονται διαδικασίες με το χαρακτηριστικό να μην έχουν μνήμη, που έχουν επιπλέον περιορισμούς όσων αφορά τον πεπερασμένο χώρο καταστάσεων τους και τα διακριτά χρονικά διαστήματα τους [15]. Οι αλυσίδες Markov είναι ένα πολύ χρήσιμο εργαλείο μοντελοποίησης και βρίσκουν εφαρμογές σε προβλήματα όπως ταξινόμησης και αναζήτησης, ωστόσο παρουσιάζει σημαντικά προβλήματα που η κατάσταση του μοντέλου δεν είναι εφικτό να αναπαρασταθεί από αυτές και για αυτόν τον λόγο αναπτύχθηκε ένα ισχυρότερο μοντέλο αυτό των κρυπτομαρκοβιανών μοντέλων.

3.2.2 Κρυπτομαρκοβιανά Μοντέλα

Ένα κρυπτομαρκοβιανό μοντέλο (hidden Markov model-HMM) είναι μια ξεχωριστή περίπτωση μιας μαρκοβιανής διαδικασίας στην οποία η λειτουργία του συστήματος αποτελεί ένα μαύρο κουτί για τον παρατηρητή δηλαδή είναι κρυμμένη. Υπάρχει διαφάνεια μόνο στις τιμές εξόδου οι οποίες προκύπτουν όμως από διαδικασίες μη φανερές. Παρόλα αυτά όπως και με τις αλυσίδες Markov η διαδικασία μπορεί να αναπαρασταθεί από ένα κατευθυνόμενο γράφημα, όπου κάθε κόμβος αντιστοιχεί σε μια κατάσταση από ένα ορισμένο σύνολο S και όπου κάθε άκρο αντιστοιχεί σε μια μετάβαση. Ακόμα υπάρχει μια μεταβλητή X_t η οποία παίρνει τιμές από το S , με το t να υποδηλώνει την τιμή της X την χρονική στιγμή t . Ωστόσο, η συγκεκριμένη μεταβλητή

είναι κρυμμένη στον παρατηρητή, αντίθετα με μια τυχαία μεταβλητή Y , η οποία αποθηκεύει τις τιμές των εξόδων του HMM, με την Y_t να υποδηλώνει την τιμή την μεταβλητής Y την χρονική στιγμή t . Η συνάρτηση $f(X_t)$, από την οποία προκύπτει και η τιμή της Y μπορεί να έχει οποιαδήποτε μορφή και πολυπλοκότητα. Παρακάτω, ένα χαρακτηριστικό γραφικό παράδειγμα HMM (εικόνα 3.7) βοηθάει στην κατανόηση των μοντέλων αυτών.



Εικόνα 3.7: Κομμάτι από γράφημα HMM. Στην πάνω σειρά οι καταστάσεις του μοντέλου κάθε στιγμή με τις αντίστοιχες εξόδους τους στην κάτω σειρά.

Στην παραπάνω εικόνα, οι καταστάσεις $x_{t-1}, x_t, x_{t+1}, \dots$ παράγονται από μια αλυσίδα Markov, αλλά ο παρατηρητής έχει την δυνατότητα να παρακολουθεί μόνο τις τιμές $y_{t-1}, y_t, y_{t+1}, \dots$. Ακόμα αν είναι γνωστές οι καταστάσεις $X_1, X_2, X_3, \dots, X_t$ τότε ισχύει:

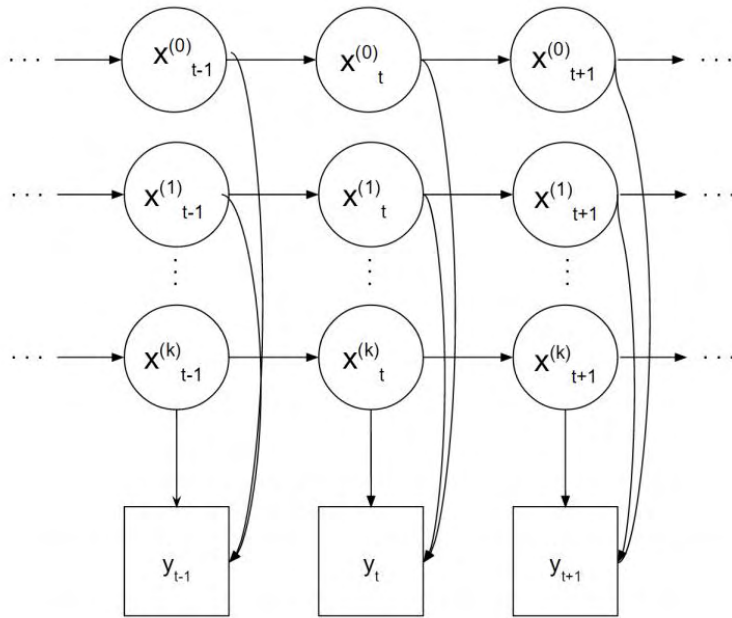
$$P(Y_t = y_t \mid X_t = x_t) = P(Y_t = y_t \mid (X_1 = x_1, X_2 = x_2, \dots, X_t = x_t))$$

Τα κρυπτομαρκοβιανά μοντέλα χρησιμοποιούνται σε μια τεράστια πληθώρα εφαρμογών πρόβλεψης και αναγνώρισης, ωστόσο παρά την ισχύ τους παρουσιάζουν αδυναμίες στην επίλυση μερικών προβλημάτων. Για αυτό τον λόγο, αναπτύχθηκαν τα παραγοντικά κρυπτομαρκοβιανά μοντέλα, τα οποία είναι συνδυασμός πολλών HMM [16].

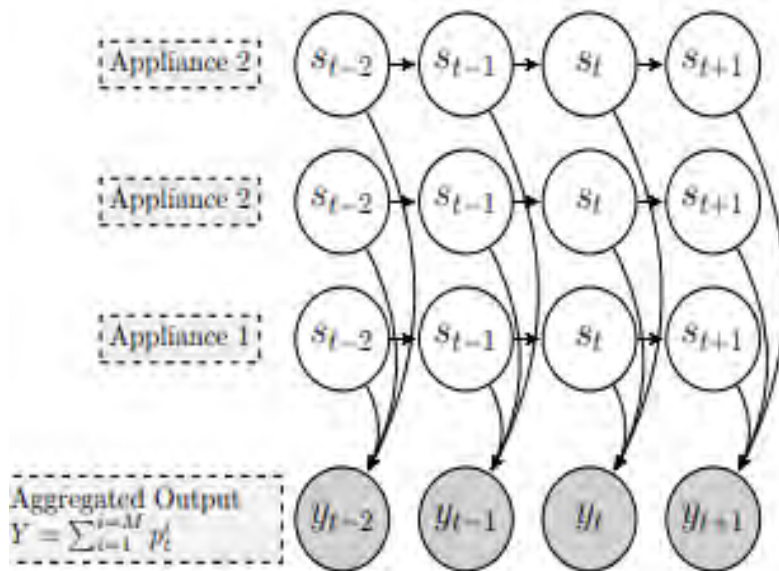
3.2.3 Παραγοντικά Κρυπτομαρκοβιανά Μοντέλα

Τα HMM παρόλο που είναι πολύ χρήσιμα για την επίλυση μεγάλης ποικιλίας προβλημάτων υστερεί σε θέμα χωρητικότητας. Χαρακτηριστικό παράδειγμα είναι οι εφαρμογές αναγνώρισης φορτίου οικιακών συσκευών, όπου το πλήθος των συσκευών ενός νοικοκυριού θα ήταν αδύνατων να αναπαρασταθεί. Την λύση σε τέτοιου είδους προβλήματα δίνουν τα παραγοντικά κρυπτομαρκοβιανά μοντέλα (Factorial hidden Markov models-FHMMs) τα οποία είναι μια γενική μορφή των HMM και διαθέτουν περισσότερες από μια μεταβλητές για την κωδικοποίηση των καταστάσεων των ζητούμενων ποσοτήτων. Συγκεκριμένα, ενώ το HMM έχει μόνο μια μεταβλητή X , τα FHMM παρέχουν k διαφορετικές μεταβλητές $X^{(0)}, X^{(1)}, \dots, X^{(k)}$ που αντιπροσωπεύουν καταστάσεις. Εύκολα βγαίνει το συμπέρασμα, ότι ένα FHMM με k μοντέλα και n καταστάσεις ισοδυναμεί με n^k καταστάσεις.

Ένα FHMM μπορεί να αναπαρασταθεί από έναν συνδυασμό k HMM που εξελίσσονται ταυτόχρονα και ανεξάρτητα από την πάροδο του χρόνου, με την μεταβλητή $X^{(i)}$, να αναπαριστά στην κατάσταση του i -στού μοντέλου (Εικόνα 3.8). Έπειτα, σε κάθε χρονικό βήμα υπάρχει μια αλληλουχία μεταβλητών για κάθε HMM στο FHMM: $X_0^{(0)}, X_1^{(1)}, \dots, X_k^{(k)}$ για κάθε χρονική στιγμή t . Κάθε ξεχωριστό HMM στο μοντέλο έχει τον δικό του αριθμό καταστάσεων n_i , τις δικές του συναρτήσεις $f_j^{(i)}$, με το i να αναφέρεται στο i -στο HMM και το j στην j -οστή κατάσταση. Επιπλέον, κάθε HMM στο παραγοντικό μοντέλο διαθέτει την δικιά του έξοδο, η οποία όμως δεν είναι ορατή στον παρατηρητή. Η ορατή έξοδο την χρονική στιγμή t , η Y_t , είναι ένας συνδυασμός των εξόδων των HMM του μοντέλου, $X_t^{(0)}, X_t^{(1)}, \dots, X_t^{(k)}$ [17]. Ένα γενικευμένο παράδειγμα ενός FHMM απεικονίζεται παρακάτω (Εικόνα 3.9):



Εικόνα 3.8: Αναπαράσταση ενός FHMM. Οι τρεις πρώτες σειρές αντιστοιχούν στις τρεις ανεξάρτητες αλυσίδες Markov.



Εικόνα 3.9: Χρήση FHMM για επιμερισμό της ενέργειας.

3.3 Αλγόριθμοι Αποσύνθεσης

3.3.1 DDSC

Ο αλγόριθμος διακριτής αποσύνθεσης αραιής κωδικοποίησης (Discriminative Disaggregation Sparse Coding-DDSC) πρωτοπαρουσιάστηκε από τους Kolter, Batra και Ng το 2010. Πρόκειται για έναν αλγόριθμο βασισμένο στην τεχνική της αραιής κωδικοποίησης, η οποία είναι μια μέθοδος μη-επιτηρούμενης μάθησης που εκπαιδεύεται από μικρό όγκο δεδομένων, δηλαδή δουλεύει με δεδομένα τα οποία έχουν προκύψει από δειγματοληψία χαμηλής συχνότητας. Ο DDSC είναι αλγόριθμος σχεδιασμένος να για περιβάλλοντα με χαμηλές συχνότητες επομένως θα ήταν δύσκολο να συγκριθεί με άλλους αλγορίθμους, παρόλα αυτά αποτελεί μια εναλλακτική λύση όταν τα δεδομένα συλλέγονται από ωριαίες μετρήσεις.

Αραιή κωδικοποίηση έχουμε όταν ενός δεδομένου διανύσματος x , προσπαθούμε να το εκφράσουμε ως άθροισμα k βάσεων θ :

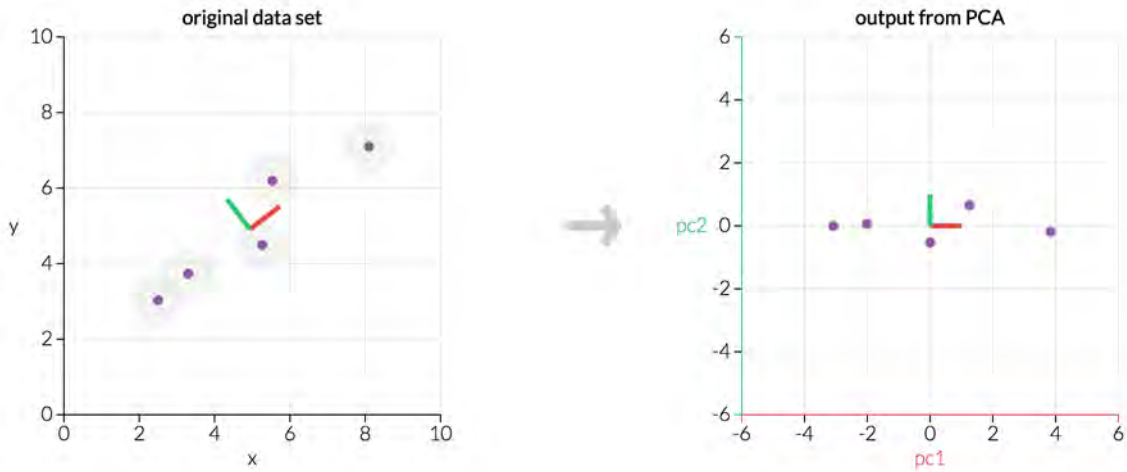
$$x = \sum_{i=1}^k a_i \theta_i$$

όπου το a_i είναι ο i -στός συντελεστής της βάσης θ_i . Εδώ βλέπουμε ομοιότητες με την μέθοδο Principal Component Analysis (PCA), η οποία χρησιμοποιείται για την μείωση ενός μεγάλου συνόλου μεταβλητών σε ένα μικρότερο το οποίο όμως περιέχει τις περισσότερες πληροφορίες του πρώτου. Ουσιαστικά, μετατρέπει τα δεδομένα σε ένα νέο σύστημα με συντεταγμένες με τέτοιο τρόπο ώστε οι βάσεις αυτού του συστήματος (principal components) να είναι ευθυγραμμισμένες με τις κατευθύνσεις στις οποίες το σύνολο των δεδομένων παρουσιάζει μεγαλύτερη μεταβλητότητα (Εικόνα 3.10). Συνεπώς θα μπορούσαμε να πούμε ότι η πρώτη βάση (principal components) αναπαριστά το υψηλότερο ποσοστό των δεδομένων, η δεύτερη αναπαριστά το δεύτερο υψηλότερο κλπ. Στην περίπτωση που οι αρχικές βάσεις αναπαριστούν επαρκώς

δηλαδή ένα μεγάλο ποσοστό των δεδομένων, οι υπόλοιπες μπορούν να αγνοηθούν χωρίς να έχουν επίδραση στην αναπαράσταση της εισόδου. Η διαφορά του PCA με την αραιή κωδικοποίηση είναι η εξής: αν ο PCA μπορεί να αναπαραστήσει ένα διάνυσμα διάστασης n με k βάσεις, με $n > k$, η αραιή κωδικοποίηση προσπαθεί να το αναπαραστήσει με $k > n$. Ωστόσο, η εκμάθηση με υπερβολικά πολλές βάσεις οδηγεί στο πρόβλημα ότι το σύνολο των συντελεστών $\alpha_1, \alpha_2, \dots, \alpha_k$ δεν προσδιορίζεται πλέον μοναδικά από την είσοδο και τις βάσεις. Προκειμένου να λυθεί αυτό η μέθοδος της αραιής κωδικοποίησης εφαρμόζει ένα κριτήριο με το οποίο οι συντελεστές που απέχουν αρκετά από το μηδέν διαχειρίζονται διαφορετικά. Το πρόβλημα της αραιής κωδικοποίησης ενός συνόλου m διανυσμάτων ορίζεται πλήρως ως:

$$\underset{a_i^{(j)}, \theta_i}{\text{minimize}} \sum_{j=1}^m \left\| x^{(j)} - \sum_{i=1}^k a_i^{(j)} \theta_i \right\| + \lambda \sum_{i=1}^k S(a_i^{(j)})$$

όπου το $x^{(j)}$ είναι το j -στο διάνυσμα στο σύνολο, $\alpha^{(j)}$ είναι το διάνυσμα των συντελεστών το οποίο αναπαριστά το j -στο διάνυσμα στο σύνολο, θ_i είναι η i -στη βάση και $a_i^{(j)}$ είναι ο i -στος συντελεστής ο οποίος χρησιμοποιείται για την αναπαράσταση του j διανύσματος. $S(\alpha)$ είναι μια συνάρτηση η οποία επιστρέφει ένα κόστος βασισμένο στην σπανιότητα του α και λ είναι μια παράμετρος η οποία χρησιμοποιείται για να μεταβάλλει τον βαθμό στον εφαρμόζεται το κριτήριο. Εδώ έχουμε να κάνουμε με πρόβλημα βελτιστοποίησης το οποίο χρειάζεται να βελτιστοποιείται ταυτόχρονα σε δύο διαστάσεις. Η μία αναζητεί τον κατάλληλο συντελεστή για κάθε διάνυσμα του συνόλου της εισόδου και η άλλη αναζητεί το κατάλληλο σύνολο των βάσεων, το οποίο είναι κοινό για όλα τα διανύσματα στο σύνολο.



Εικόνα 3.10: Μετατροπή συνόλου δεδομένων με την μέθοδο PCA.

Η DDSC είναι μια μέθοδος μηχανικής μάθησης υπο-επίβλεψη. Ας θεωρήσουμε ως m τον αριθμό των σπιτιών, T τον αριθμό των σημείων από όπου παίρνουμε δεδομένα για κάθε σπίτι που προέρχονται από δειγματοληψία χαμηλής συχνότητας και τον αριθμό των διαφορετικών κλάσεων συσκευών που θέλουμε να αποσυνθέσουμε. Η είσοδος του DDSC είναι ένας πίνακας $T \times m$ με δεδομένα ενέργειας, όπου κάθε στήλη περιέχει τα δεδομένα ενέργειας για κάθε σπίτι. Η είσοδος ακόμα περιέχει τα X_1, X_2, \dots, X_k , τα οποία είναι επίσης $T \times m$ πίνακες που αναπαριστούν τα ξεχωριστά σήματα για κάθε κλάση συσκευής. Ο DDSC εφαρμόζει την αραιή κωδικοποίηση σε όλους τους πίνακες X_i με στόχο την προσέγγιση της $X_i = B_i A_i$ όπου το B_i είναι οι βάσεις και A_i οι συντελεστές. Όταν τα παραπάνω μαθευτούν για κάθε κλάση, η αποσύνθεση ενός σήματος X γίνεται ζήτημα προβλήματος βελτιστοποίησης:

$$\hat{A}_{1..k} = \arg \min_{A_{1..k} \geq 0} \left\| \bar{X} - B_{1..k} A_{1..k} \right\|^2 + \lambda \sum_{i,p,q} (A_i)_{pq} \equiv F(\bar{X}, B_{1..k}, A_{1..k})$$

όπου το A_{pq} είναι το στοιχείο στην p σειρά και στην q στήλη του A . Έπειτα τα A_1, A_2, \dots, A_k χρησιμοποιούνται για την αποσύνθεση του σήματος X με τον ακόλουθο τρόπο:

$$\hat{X}_i = B_i \hat{A}_i,$$

Όπου το X_i είναι το σήμα το οποίο αντιστοιχεί στην i των συσκευών. Ο αλγόριθμος φαίνεται ολοκληρωμένος παρακάτω [18]:

Algorithm 1 Discriminative disaggregation sparse coding

Input: data points for each individual source $X_i \in \mathbb{R}^{T \times m}$, $i = 1, \dots, k$, regularization parameter $\lambda \in \mathbb{R}_+$, gradient step size $\alpha \in \mathbb{R}_+$.

Sparse coding pre-training:

1. Initialize B_i and A_i with positive values and scale columns of B_i such that $\|b_i^{(j)}\|_2 = 1$.
2. For each $i = 1, \dots, k$, iterate until convergence:
 - (a) $A_i \leftarrow \arg \min_{A \geq 0} \|X_i - B_i A\|_F^2 + \lambda \sum_{p,q} A_{pq}$
 - (b) $B_i \leftarrow \arg \min_{B \geq 0, \|b^{(j)}\|_2 \leq 1} \|X_i - B A_i\|_F^2$

Discriminative disaggregation training:

3. Set $A_{1:k}^* \leftarrow A_{1:k}$, $\tilde{B}_{1:k} \leftarrow B_{1:k}$.
4. Iterate until convergence:
 - (a) $\hat{A}_{1:k} \leftarrow \arg \min_{A_{1:k} \geq 0} F(\bar{X}, \tilde{B}_{1:k}, A_{1:k})$
 - (b) $\tilde{B} \leftarrow \left[\tilde{B} - \alpha \left((\bar{X} - \tilde{B} \hat{A}) \hat{A}^T - (\bar{X} - \tilde{B} A^*) (A^*)^T \right) \right]_+$
 - (c) For all i, j , $b_i^{(j)} \leftarrow b_i^{(j)} / \|b_i^{(j)}\|_2$.

Given aggregated test examples \bar{X}' :

5. $\hat{A}'_{1:k} \leftarrow \arg \min_{A_{1:k} \geq 0} F(\bar{X}', \tilde{B}_{1:k}, A_{1:k})$
 6. Predict $\hat{X}'_i = B_i \hat{A}'_i$.
-

3.3.2 CO

Η αναγνώριση φορτίου μέσω μη παρεμβατικής τεχνικής μπορεί να θεωρηθεί ως ένα πρόβλημα συνδυαστικής βελτιστοποίησης (Combinatorial Optimization-CO) που περιγράφεται ως εξής:

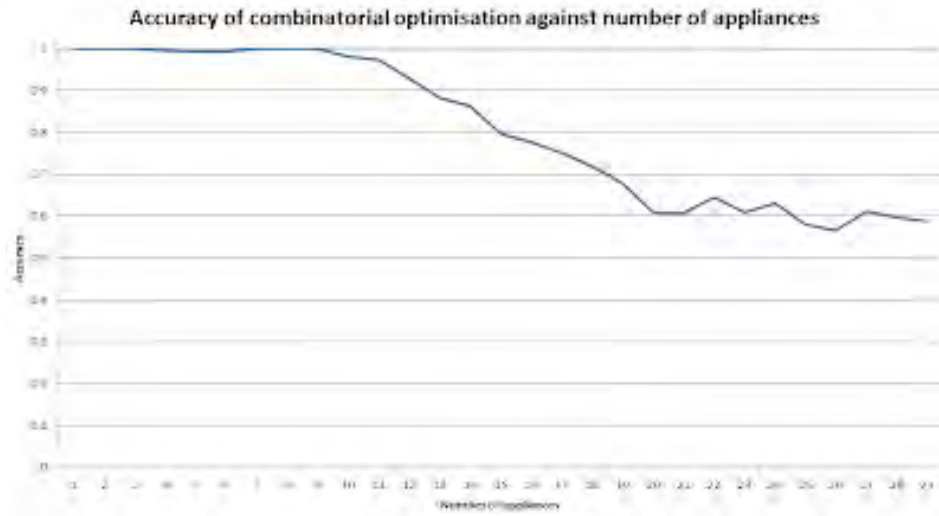
Για κάθε χρονική στιγμή ελαχιστοποιούμε την τιμή της συνάρτησης f :

$$f = \sum_{n=1}^N |P_{agg} - (I_n \times P_n)|$$

όπου P_{agg} είναι η οικιακή συνολική ενέργεια, N ο αριθμός των συσκευών, I το διάνυσμα που περιέχει δυαδική πληροφορία σχετικά με την λειτουργία των συσκευών (τιμή 1 αν η συσκευή είναι σε λειτουργία, διαφορετικά τιμή 0) και P το διάνυσμα που περιέχει τις τιμές της καταναλισκόμενης ενέργειας κάθε συσκευής όταν αυτές είναι σε λειτουργία. Η προσέγγιση αυτή προτάθηκε από τον Hart ο οποίος ισχυρίστηκε ότι πρόκειται για ένα NP (nondeterministic polynomial)-πρόβλημα, πρόβλημα του οποίου την λύση μπορούμε να μαντέψουμε και να επαληθεύσουμε σε πολυωνυμικό χρόνο χωρίς να ακολουθήσουμε κάποιον συγκεκριμένο κανόνα για να προβούμε στην εικασία. Παράλληλα, ο Hart το έθεσε και ως ένα πρόβλημα σταθμισμένου συνόλου και ανέφερε ότι ο αριθμός των συνδυασμών των συσκευών σε κατάσταση λειτουργίας κλιμακώνεται εκθετικά με τον αριθμό των συσκευών. Ωστόσο, παρά την υπολογιστική δυσκολία του, η βέλτιστη λύση μπορεί να βρεθεί μέσα σε δευτερόλεπτα για έναν ρεαλιστικό αριθμό συσκευών.

Στο σημείο αυτό ας δούμε μια προσπάθεια διερεύνησης της απόδοσης της παραπάνω τεχνικής χρησιμοποιώντας προσομοιωμένα δεδομένα. Τα δεδομένα δημιουργήθηκαν έπειτα από επιτήρηση και καταγραφής της καταναλισκόμενης ενέργειας 30 συσκευών, με ρεαλιστικές απαιτήσεις ισχύος και χρήσης, σε διάστημα 24 ωρών. Η πραγματική συνολική καταναλισκόμενη ενέργεια του νοικοκυριού υπολογίστηκε αθροίζοντας την ενέργεια κάθε συσκευής την δεδομένη χρονική στιγμή. Ωστόσο, πρόκειται για ένα εξειδικευμένο παράδειγμα καθώς γίνεται χωρίς την παρουσία θορύβου τόσο σε επίπεδο συσκευών όσο και νοικοκυριών.

Σύμφωνα με τα αποτελέσματα, καταλήγουμε στο συμπέρασμα ότι η απόδοση αυτής της τεχνικής μειώνεται καθώς ο αριθμός των συσκευών του νοικοκυριού αυξάνεται. Αναλυτικότερα, στο παρακάτω γράφημα (Εικόνα 3.11), φαίνεται ακρίβεια της τεχνικής για κάθε αριθμό συσκευών.



Εικόνα 3.11: Η μέθοδος βελτιστοποίησης εφαρμόστηκε σε διαστήματα 1 ώρας έως 24 ώρες δεδομένων και η ακρίβεια υπολογίστηκε ως το μέσο κλάσμα των σωστών ταξινομήσεων.

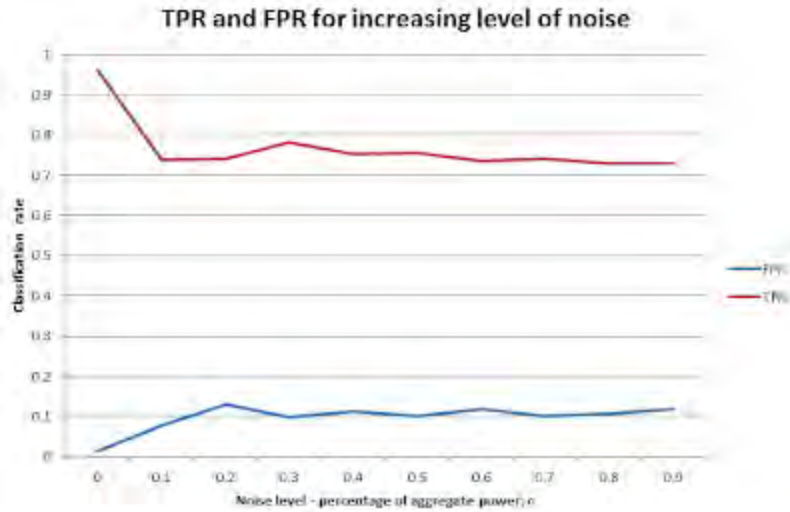
Γίνεται εύκολα αντιληπτό ότι αυτή η τεχνική είναι πολύ αποτελεσματική για μικρό σύνολο συσκευών ($N < 10$) ενώ όσο αυξάνει ο αριθμός των συσκευών, μειώνεται δραστικά η ακρίβειά της και σταθεροποιείται σε μια τιμή μεταξύ του 0.5 και 0.6. Μια καλή ιδέα λοιπόν, δεδομένης της ικανοποιητικής ακρίβειας της μεθόδους αυτής σε λίγες συσκευές, θα ήταν η επιλογή των πιο ενεργοβόρων συσκευών προς επιτήρηση.

Για την διερεύνηση της ακρίβειας και πως αυτή επηρεάζεται από την παρουσία θορύβου στα δεδομένα, προστέθηκε θόρυβος Gauss στις τιμές ανάγνωσης ισχύος της προκαθορισμένης μορφής:

$$N(x|P_{agg}, (P_{agg}c)^2)$$

όπου N η γκαουσιανή κατανομή, x το προκαθορισμένο επίπεδο συνολικού θορύβου, P_{agg} το πραγματικό προκαθορισμένο επίπεδο συνολικού θορύβου και c το επίπεδο θορύβου.

Στο παρακάτω γράφημα (Εικόνα 3.12) καταγράφεται η ευαισθησία (TPR) και η ιδιαιτερότητα (FPR) ανάλογα με την αύξηση του θορύβου.



Εικόνα 3.12: Αμέσως μετά την εισαγωγή κάποιο θορύβου, το TPR πέφτει σε περίπου 0.7. Συνεπώς η προσέγγιση αυτή δεν είναι αποδοτική για ακόμη και μικρά επίπεδα θορύβου στις μετρήσεις συνολικής ισχύος.

Με στόχο την μεγαλύτερη ακρίβεια της παραπάνω προσέγγισης μπορούν να γίνουν δυο επεκτάσεις. Πρώτα μπορούμε να τροποποιήσουμε την υπόθεση ότι μια συσκευή μπορεί να έχει μια κατάσταση κατά την οποία καταναλώνει ρεύμα και να υποθέσουμε ότι μπορεί να λειτουργεί σε μια από K καταστάσεις, με κάθε μια από αυτές να έχει μια συγκεκριμένη κατανάλωση. Έτσι έχουμε λοιπόν:

$$I_{nk} \in \{0,1\}$$

και

$$\sum_{k=1}^{k=K} I_{nk} = 1$$

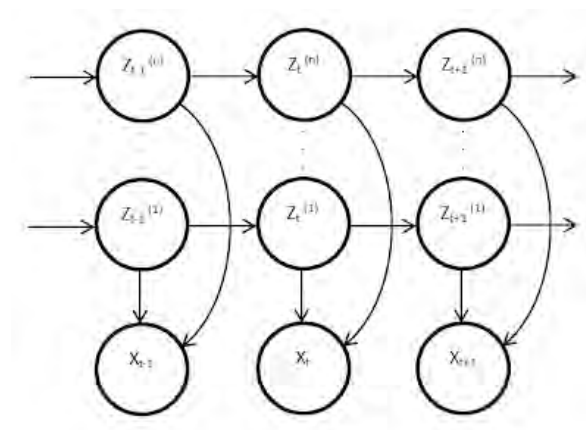
Επομένως ο βέλτιστος συνδυασμός των συσκευών θα δίνεται από την ελαχιστοποίηση της τιμής της συνάρτησης f σε κάθε χρονική στιγμή:

$$f = \sum_{n=1}^N \left| P_{agg} - \sum_{k=1}^{k=K} I_{nk} P_{nk} \right|$$

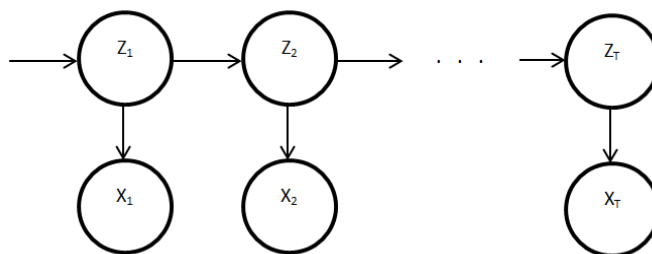
Ακόμα θα μπορούσε να αλλάξει η υπόθεση ότι οι καταστάσεις της συσκευής είναι ανεξάρτητες μεταξύ τους, πράγμα το οποίο θα μας επέτρεπε να λάβουμε επιπλέον πληροφορίες αλλά με κόστος την αύξηση της πολυπλοκότητας του προβλήματος. [19]

3.3.3 HMM

Σε αυτήν την ενότητα θα αναλύσουμε πως με την βοήθεια των μαρκοβιανών μοντέλων μπορούμε να αναπαραστήσουμε πολλαπλές συσκευές. Όπως είδαμε σε προηγούμενο κεφάλαιο, ο συνδυασμός αλυσίδων Markov οι οποίες αναπαριστούν καταστάσεις των συσκευών και μια μεταβλητή που θα αντιπροσωπεύει την μετρούμενη ισχύ της μπορούν να μοντελοποιηθούν σαν ένα κρυπτομαρκοβιανό μοντέλο:



Ένα τέτοιο μοντέλο θα περιέχει μια μονή αλυσίδα μεταβλητών, κάθε μια από τις οποίες θα αποκτούν τόσες διακριτές τιμές όσες και ο αριθμός των συνδυασμών των καταστάσεων κάθε συσκευής. Για παράδειγμα, ένα παραγοντικό κρυπτομαρκοβιανό μοντέλο (FHMM) με τρεις αλυσίδες των δύο, τριών και τεσσάρων καταστάσεων αντίστοιχα, θα ισοδυναμεί με ένα απλό κρυπτομαρκοβιανό μοντέλο (HMM) με μια αλυσίδα των εικοσετεσσάρων καταστάσεων:



Τώρα ας διακρίνουμε τις παραμέτρους που περιγράφουν το FHMM: θέτουμε p_i την πιθανότητα κατάστασης μιας συσκευής στο πρώτο χρονικό διάστημα, A τον πίνακα πιθανότητας αλλαγής κατάστασης και p_{hi} την πιθανότητα πυκνότητας (probability density) της i κρυφής μεταβλητής. Θεωρώντας ότι έχουμε στην διάθεσή μας δεδομένα για κάθε συσκευή, οι παράμετροι του μοντέλου μπορούν να μαθευτούν με την χρήση του αλγορίθμου Expectation-Maximisation, που είναι ένας αλγόριθμος για την εκτίμηση της μέγιστης πιθανότητας σε μοντέλα με κρυφές μεταβλητές. Έπειτα ακολουθεί ο συνδυασμός των παραμέτρων των μοντέλων κάθε συσκευής με αποτέλεσμα ένα σύνολο παραμέτρων που περιγράφουν ένα συνδυαστικό κρυπτομαρκοβιανό μοντέλο. Οι τελικές παράμετροι υπολογίζονται ως εξής: p_i το αποτέλεσμα του συνδυασμού κάθε ξεχωριστής αρχικής πιθανότητας, A το αποτέλεσμα του συνδυασμού των πιθανοτήτων αλλαγής κατάστασης και p_{hi} το άθροισμα κάθε συνδυασμού των αρχικών p_{hi} . Επόμενο βήμα είναι η εύρεση της πιο πιθανής αλυσίδας των κρυφών μεταβλητών του μοντέλου. Εδώ το πρόβλημα ανάγεται σε πρόβλημα δυναμικού προγραμματισμού όπου καλούμαστε να βρούμε την πιθανότερη ακολουθία καταστάσεων και την λύση έρχεται να δώσει ο αλγόριθμος Viterbi [20]. Μια σύντομη περιγραφή του συγκεκριμένου αλγορίθμου είναι η εξής: δεδομένης μιας ακολουθίας παρατηρούμενων συμβόλων y_1, \dots, y_T T στοιχείων με καταστάσεις στο πεπερασμένο σύνολο S , π_i η αρχική πιθανότητα της κατάστασης i και $a_{i,j}$ οι πιθανότητες μετάβασης από την κατάσταση i στην κατάσταση j . Η πιθανότερη ακολουθία x_1, x_2, \dots, x_T από την οποία προέκυψαν οι παρατηρήσεις υπολογίζεται από τις σχέσεις:

$$V_{1,k} = P(y_1 | k) \cdot \pi_k$$

$$V_{t,k} = \max_{x \in S} (P(y_t | k) \cdot a_{x,k} \cdot V_{t-1,x})$$

όπου $V_{t,k}$ είναι η πιθανότητα της πιο πιθανής ακολουθίας καταστάσεων για τις πρώτες t παρατηρήσεις που καταλήγει στην κατάσταση k . Ωστόσο, οι σχέσεις αυτές δεν δίνουν την πιθανότερη ακολουθία καταστάσεων, αλλά αυτό γίνεται σε ένα δεύτερο βήμα όπου ο αλγόριθμος υπολογίζει το μονοπάτι προς τα πίσω [21]:

$$x_T = \arg \max_{x \in S} (V_{T,x})$$

$$x_{t-1} = \text{Ptr}(x_t, t)$$

3.3.4 MLE

Η εκτίμηση μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation-MLE) είναι μια μέθοδος που καθορίζει τιμές των παραμέτρων ενός μοντέλου. Οι τιμές των παραμέτρων βρίσκονται έτσι ώστε να μεγιστοποιούν την πιθανότητα ότι η διαδικασία που περιγράφεται από το μοντέλο παρήγαγε τα δεδομένα που πράγματι παρατηρήθηκαν. Στην δικιά μας περίπτωση, ο αλγόριθμος αποσυνθέτει μια μοναδική συσκευή με βάση τα χαρακτηριστικά της και χρησιμοποιεί τη μέγιστη πιθανότητα όλων των χαρακτηριστικών.

Υπάρχουν τρία χαρακτηριστικά υπό εξέταση: η λειτουργία ON, η λειτουργία OFF και η διάρκεια μεταξύ αυτών των δύο. Τα προαναφερθέντα χαρακτηριστικά στην συνέχεια διαμορφώνονται με τεχνικές όπως Gaussian, Gaussian Mixtures ή Poisson. Για κάθε ζεύγος γεγονότων που ανιχνεύεται, ο αλγόριθμος θα εξάγει αυτά τα τρία χαρακτηριστικά και θα αξιολογήσει την μέγιστη πιθανότητα για το συγκεκριμένο ζεύγος γεγονότων να ανήκει σε μια συγκεκριμένη συσκευή.

ΚΕΦΑΛΑΙΟ 4

ΔΕΔΟΜΕΝΑ & ΕΡΓΑΛΕΙΑ

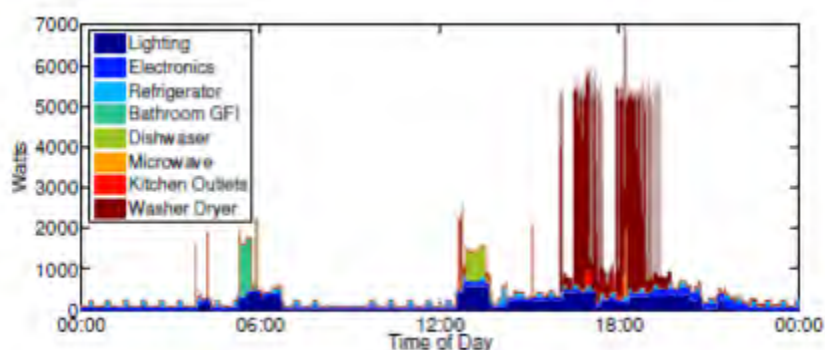
4.1 Σύνολα Δεδομένων

Για την εκτέλεση πειραμάτων με σκοπό την αξιολόγηση των παραπάνω αλγορίθμων απαραίτητη κρίνεται η εύρεση κατάλληλων δεδομένων. Τα δεδομένα αυτά θα αποτελέσουν την πηγή για την μάθηση των αλγορίθμων αλλά και για την αξιολόγηση της αποτελεσματικότητας του συστήματος της μη παρεμβατικής παρακολούθησης φορτίου. Το περιεχόμενο των δεδομένων περιγράφει την κατάσταση και τις λειτουργίες των ηλεκτρικών συσκευών και συλλέγονται από πραγματικά νοικοκυριά που λειτουργούν ανεπηρέαστα μέσω ειδικών μετρητών στα σημεία τροφοδοσίας των συσκευών. Οι συγκεκριμένοι μετρητές καταγράφουν μεγέθη όπως είναι η ένταση του ρεύματος και η ισχύς της συσκευής αλλά υπάρχει και μετρητής που είναι υπεύθυνος για την καταγραφή της συνολικής καταναλισκόμενης ενέργειας του σπιτιού που από το σήμα της θα γίνει και η αποσύνθεσή της σε επιμέρους ενέργειες. Τα δεδομένα δειγματολειτουργούν με συγκεκριμένη συχνότητα και σχηματίζονται χρονοσειρές και οι χρονοσειρές όλων των συσκευών αποτελούν το σύνολο των δεδομένων. Ωστόσο, θεωρείται σημαντικό να συγκρίνουμε και να κρίνουμε τους αλγορίθμους αποσύνθεσης σε σύνολα δεδομένων που συλλέχτηκαν από διαφορετικά νοικοκυριά διαφορετικών χωρών ώστε να υπάρχει μια ποικιλομορφία στον περιεχόμενό τους λόγω της αλλαγής στο προφίλ χρήσης των συσκευών. Τέτοια σύνολα δεδομένων διατίθενται ελεύθερα ή τουλάχιστον τα περισσότερα για την εκτέλεση πειραμάτων και μερικά από αυτά θα περιγράψουμε παρακάτω καθώς είναι απαραίτητο να κατανοηθούν τα χαρακτηριστικά τους.

4.1.1 REDD

Από τα πρώτα σύνολα δεδομένων που αναπτύχθηκαν είναι αυτό που ονομάζεται Σύνολο Δεδομένων Αναφοράς Ενεργειακής Κατανομής - Reference Energy Disaggregation Data Set (REDD). Δημοσιεύθηκε το 2011 από τους J. Zico Kolter

Matthew και J. Johnson του MIT. Τα δεδομένα αυτά πάρθηκαν από την καταγραφή έξι διαφορετικών οικιών στις Η.Π.Α που κάθε μια από αυτές είχε από 10 μέχρι 20 σημεία μέτρησης. Το χρονικό διάστημα των μετρήσεων ήταν μερική μήνες, με περίοδο τριών δευτερολέπτων και δίνουν πληροφορίες συσκευών που συναντάει κανείς σε ένα νοικοκυριό. Για κάθε κατοικία, υπάρχει καταγεγραμμένο το ηλεκτρικό σήμα της οικιακής χρήσης σε υψηλή συχνότητα (15kHz), 2 έως 24 επιμέρους κυκλώματα, κάθε ένα συνοδευόμενο από μια ετικέτα που περιγράφει την κατηγορία της συσκευής ή πολλών συσκευών που ομαδοποιούνται σε ένα κύκλωμα. Ένα στιγμιότυπο αυτού του τύπου δεδομένων παρουσιάζεται παρακάτω (εικόνα 4.1). Το REDD αντιπροσωπεύει ένα από τα μεγαλύτερα σύνολα που είναι ελεύθερα διαθέσιμα για αποσύνθεση ενέργειας με πραγματικά φορτία [22].



Εικόνα 4.1: Παράδειγμα κατανάλωσης ενέργειας κατά την διάρκεια ενός 24ρου για μια από τις οικίες του REDD.

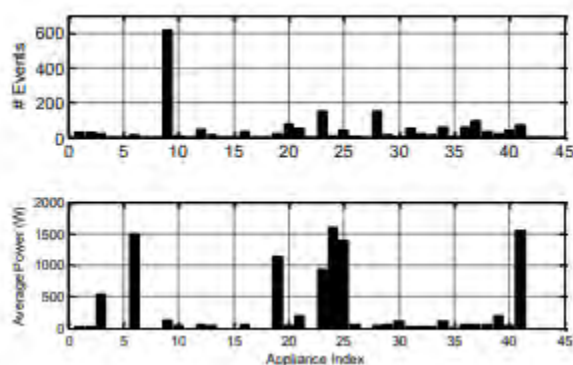
4.1.2 UK-DALE

Την χρονιά του 2015 οι J. Kelly και W. Knottenbelt δημοσίευσαν το σύνολο δεδομένων του UK-DALE το οποίο παρουσιάζει αρκετές ομοιότητες με αυτό του REDD αλλά διαθέτει πολύ μεγαλύτερες χρονοσειρές. Είναι ένα σύνολο δεδομένων ανοικτής πρόσβασης από την Μ.Βρετανία που έχει καταγεγραμμένη την κατανάλωση ηλεκτρικής ενέργειας σε επίπεδο συσκευής με ρυθμό δειγματοληψίας 16 kHz για ολόκληρο το σπίτι και 1/6 Hz για μεμονωμένες συσκευές. Αναφερόμαστε στο πρώτο σύνολο δεδομένων ανοικτής πρόσβασης στο Ηνωμένο Βασίλειο με αυτή την ανάλυση. Διαθέτει καταγραφές

από πέντε νοικοκυριά, ένα από τα οποία καταγράφηκε για έναν ολόκληρο χρόνο που είναι και η μεγαλύτερη διάρκεια που γνωρίζουμε για οποιαδήποτε σύνολο δεδομένων ενεργειακού περιεχομένου σε αυτό το ρυθμό δειγματοληψίας [23].

4.1.3 BLUED

Το BLUED (Building-Level fully-labeled dataset for Electricity Disaggregation) αποτελείται από μετρήσεις τάσης και ρεύματος για την κατοικία μιας οικογένειας στις Ηνωμένες Πολιτείες, δειγματοληψίας 12 kHz για μια ολόκληρη εβδομάδα. Πρόκειται για ένα σύνολο δεδομένων με ετικέτες για την λειτουργία συγκεκριμένων συσκευών που περιγράφει την χρήση ηλεκτρικού ρεύματος μιας οικίας για μια εβδομάδα. Περιέχει πληροφορία για τις μεταβιβάσεις των καταστάσεων των συσκευών και των δραστηριοτήτων τους (Εικόνα 4.2). Αυτό το σύνολο δεδομένων μπορεί να αξιοποιηθεί όχι μόνο από πειράματα για την αξιολόγηση της αξιοπιστίας συστημάτων μη παρεμβατικής παρακολούθησης φορτίου αλλά και από εφαρμογές ασφάλειας διαχείρισης ενέργειας, διάγνωσης σφαλμάτων και ανίχνευσης ανωμαλιών [24].



Εικόνα 4.2: Πάνω: Ιστορικό του αριθμού των συμβάντων για κάθε μια συσκευή. Κάτω: Μέση κατανάλωση ενέργειας για κάθε μια συσκευή.

4.1.4 ECO

Σε αυτό το σύνολο, το οποίο συγκεντρώθηκε από ερευνητές του ΕΤΗ της Ζυρίχης, περιέχει πληροφορία για την κατανάλωση ενέργειας έξι κατοικιών, ρυθμού

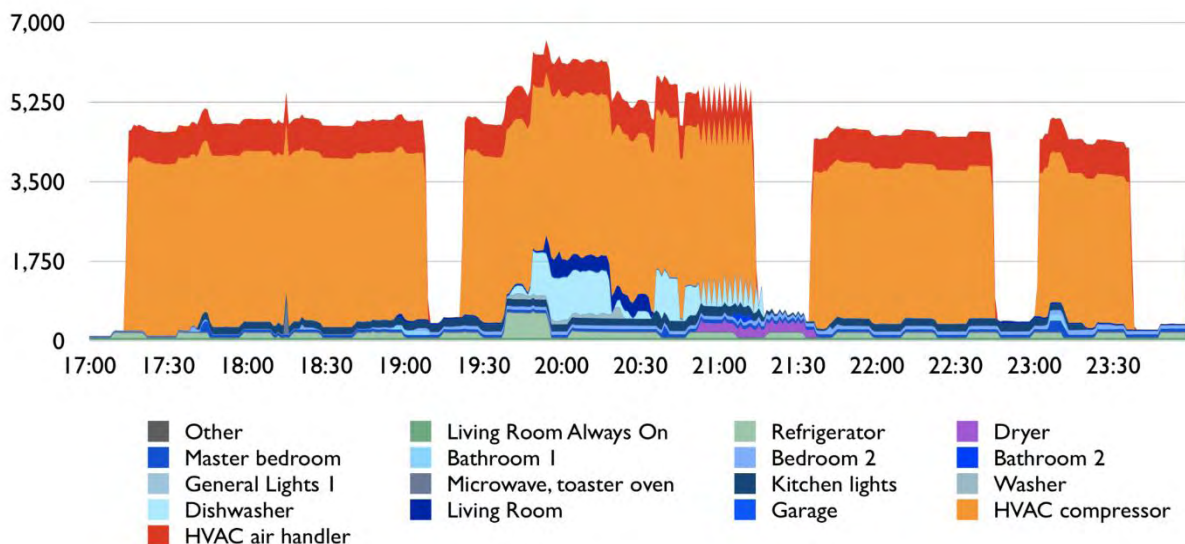
δειγματοληψίας 1 Hz διαστήματος οκτώ μηνών. Πρόκειται για ένα πρωτοποριακό σύνολο δεδομένων επειδή περιέχει όγκο δεδομένων με αρκετά λεπτομερείς μετρήσεις καταναλισκόμενης ενέργειας ακόμα και πληροφορίες για το ποιες ώρες υπήρχαν ένοικοι στις κατοικίες. Οι συσκευές οι οποίες επιτηρούνταν σε κάθε κατοικία καταγράφονται παρακάτω (εικόνα 4.3) [25].

- **Household 1:** (1) Fridge, (2) dryer, (3) coffee machine, (4) kettle, (5) washing machine, (6) PC, (7) freezer.
- **Household 2:** (1) Tablet, (2) dishwasher, (3) stove, (4) fridge, (5) TV, (6) stereo, (7) freezer, (8) kettle, (9) lamp, (10) laptops.
- **Household 3:** (1) Tablet, (2) freezer, (3) coffee machine, (4) PC, (5) fridge, (6) kettle, (7) entertainment.
- **Household 4:** (1) Fridge, (2) kitchen appliances³, (3) lamp, (4) stereo & laptop, (5) freezer, (6) tablet, (7) entertainment, (8) microwave.
- **Household 5:** (1) Tablet, (2) coffee machine, (3) kettle, (4) microwave, (5) fridge, (6) entertainment, (7) PC, router & printer, (8) fountain.
- **Household 6:** (1) Lamp, (2) laptop & printer, (3) routers, (4) coffee machine, (5) entertainment, (6) fridge, (7) kettle.

Εικόνα 4.3 Οι κατοικίες και οι συσκευές από όπου πάρθηκαν οι μετρήσεις του ECO.

4.1.5 Pecan Street

Το 2018 η εταιρία Pecan Street κυκλοφόρησε ένα πρωτοποριακό σύνολο δεδομένων που περιέχει δεδομένα κατανάλωσης ηλεκτρικού ρεύματος, τα οποία συλλέχθηκαν μέσω του εθελοντικού οικιακού ερευνητικού δικτύου (Εικόνα 4.4). Το πακέτο δεδομένων της χωρίζεται σε ένα σύνολο δεδομένων οικίας που καταγράφηκαν με ρυθμό δεκαπέντε λεπτών και ένα σύνολο δεδομένων ρυθμού ενός λεπτού 10 έως 100 νοικοκυριών στο Ώστιν χρονικής διάρκειας 365 ημερών και αφορά το επίπεδο κυκλώματος των κατοικιών.



Εικόνα 4.4: Παράδειγμα δεδομένων του πακέτου της Pecan Street αναπαριστάμενο σε διάγραμμα.

4.1.6 iAWE

Το iAWE (Indian Dataset for Ambient Water and Energy) είναι ένα σύνολο δεδομένων σχετικό με την κατανάλωση νερού και ηλεκτρικής ενέργειας ενός σπιτιού στο Νέο Δελχί της Ινδίας. Η επιτήρηση της κατανάλωσης έγινε για 73 ημέρες και η περίοδος δειγματοληψίας είναι ένα λεπτό.

4.1.7 NILM Datasets

Η παραγωγή συνόλων δεδομένων που μετρήθηκαν από πραγματικές κατοικίες μπορεί να είναι χρονοβόρα, δαπανηρή και δυνητικά ακατάλληλη για συλλογή. Οι ερευνητές για αυτό το λόγο έχουν προχωρήσει στο να δημοσιοποιούν τα σύνολα δεδομένων τους, επιτρέποντας σε άλλους ερευνητές να συγκρίνουν τις προσεγγίσεις τους με κοινά κριτήρια αναφοράς. Σήμερα είναι διαθέσιμα τα ακόλουθα σύνολα δεδομένων (Εικόνα 4.5):

Dataset	Sampling rate	Duration	# Houses / device instances	Subject	Classes	Ground truth	Country	Availability
Dataport	1 Hz to 1 minute	4+ years (actively growing)	1200+	Residential, Commercial, and Light Industrial	~70	Submeter channels	US	Free for academic use (majority) / fee for commercial
REDD	16500 Hz / 1 Hz	Several months	2 / 5	Residential	~20	Submeter channels	US	Request access
BLUED	12000 Hz	1 week	1	Residential	~30	Labelled events	US	Request access
UK-DALE	16000 Hz / 1 Hz	Up to 2 years	6 (of which 3 are at 16 kHz)	Residential	~40	Submeter channels	UK	Public FTP
PLAID	30000 Hz	5 seconds	55	Individual appliances	12	Individual appliances	US	Public
WHITED	44000 Hz	5 seconds	9	Individual appliances	46	Individual appliances	Multiple	Public
Tracebase	1 Hz	1 day	158	Individual appliances	43	Individual appliances	Germany	Request access
DRED	1 Hz / 1 min	6 months	1	Residential		Submeter channels	The Netherlands	Public
AMPds	1 minute	2 years	1	Residential	19	Submeter channels	Canada	Public
RAE	1 Hz	72 days	1	Residential	24	Submeter channels	Canada	Public
iAWE	1 Hz	73 days	1	Residential	33	Submeter channels	India	

Εικόνα 4.5: Τα διαθέσιμα σύνολα δεδομένων για την εκτέλεση πειραμάτων. Πηγή NILMwiki [26]

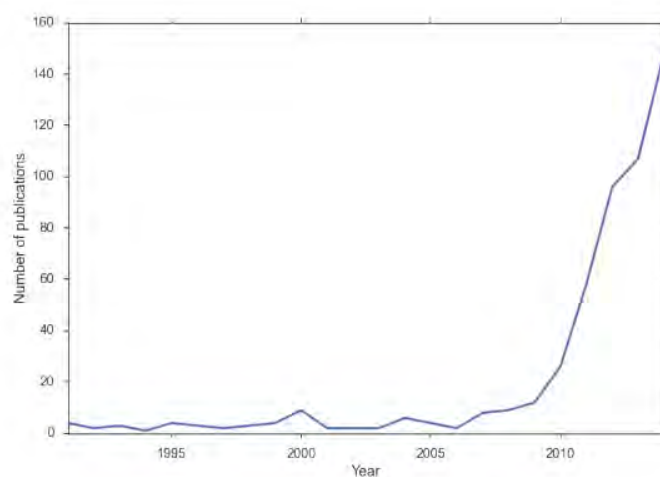
HES	2 minutes	1 year / 1 month	251	Residential	~20	Submeter channels	UK	Request access
REFIT	8 seconds	2 years	20	Residential	9	Submeter	UK	Public
ECO	1 second	8 months	6 / 45	Residential		Submeter (1 second)	Switzerland	Public
COMBED	30 seconds	1 month		academic building energy			India	
GREEND							Italy and Austria	
ACS-FI	10 seconds	2 sessions of 1 hour each	0 / 10	Home and office appliances	10	Plug-meters	Switzerland	
IHEPCDS	1 minute	4 years	1 / 3		3		France	
OCTES	7 secs	4-13 months	33 / 1	Aggregate energy consumption only	1		Finland, Iceland, Scotland	
SMART*	1 Hz	3 months	3 / 25		25		USA	
EPRI Load Shape Library	Hourly and Daily available now. 15 minute coming late-2016.						USA	
Encored TV dataset		1 month	20 / 20		1: Just TVs		Japan	To be released by early 2017?
Energy Smart Home Lab Dataset	1 Hz or 0.5 Hz		1 house	Living lab				To be released in 2017?
RBSA	15 minute	27 months	101			Submetered	Pacific Northwest, USA	

Εικόνα 4.5: Τα διαθέσιμα σύνολα δεδομένων για την εκτέλεση πειραμάτων. Πηγή NILMwiki [26]

4.2 NILMTK

4.2.1 Λόγοι για NILMTK

Τα τελευταία χρόνια το πεδίο της παρακολούθησης φορτίου μέσω μια παρεμβατικών τεχνικών έχει επεκταθεί ραγδαία λόγω του μεγάλου ενδιαφέροντος που παρουσιάζουν πολλοί οργανισμοί (Εικόνα 4.6). Ωστόσο, η σύγκριση και η αξιολόγηση αυτών των τεχνικών είναι σχεδόν αδύνατη λόγω του μεγάλου και ποικίλου όγκου δεδομένων που χρησιμοποιούνται. Κάθε σύνολο δεδομένων που έχει δημοσιευθεί χρησιμοποιεί διαφορετική μορφή και η εισαγωγή δεδομένων από διαφορετικά σύνολα απαιτεί μεγάλη τεχνική προσπάθεια. Αυτό μειώνει την ικανότητα να μετρείται η γενικευσιμότητα των αλγορίθμων. Ακόμα, η αδυναμία σύγκρισής των αλγορίθμων επιμερισμού οφείλεται και στο γεγονός ότι κάθε αλγόριθμος που παρουσιάζεται κατά καιρούς χρησιμοποιεί διαφορετικές μετρικές και διαφορετικές εισόδους. Έτσι, μπροστά σε αυτήν την πρόκληση αναπτύχθηκε από ομάδα ερευνητών το NILMTK (Non-intrusive Load Monitoring Toolkit), μια εργαλειοθήκη ανοιχτού κώδικα, σχεδιασμένη ειδικά για την αξιολόγηση αλγορίθμων αποσύνθεσης ενέργειας για ερευνητικούς σκοπούς. Συγκεκριμένα, παρέχει στον χρήστη εργαλεία μετατροπής συνόλου δεδομένων, συλλογή αλγορίθμων επιμερισμού και μετρικές ακρίβειας.



Εικόνα 4.6 : Άρθρα σχετικά με το NILM που δημοσιεύτηκαν κάθε χρόνο. Πηγή:

Parson 2015

4.2.2 Επισκόπηση του NILMTK

Το NILMTK είναι γραμμένο στην γλώσσα προγραμματισμού Python με επίπεδη μορφή αρχείων (μια επίπεδη βάση δεδομένων αρχείων αποθηκεύει τα δεδομένα σε μορφή απλού κειμένου) εισόδου και εξόδου, εκτός από τις δυαδικές μορφές υψηλής απόδοσης, εξασφαλίζοντας την απαραίτητη συμβατότητα με υπάρχοντες αλγορίθμους γραμμένους σε οποιαδήποτε γλώσσα και σχεδιασμένους για οποιαδήποτε πλατφόρμα. Το αρχείο NILMTK περιέχει κλάσεις και μεθόδους με τις οποίες επιτυγχάνονται οι λειτουργίες του. Τα στάδια από τα οποία περνάνε και επεξεργάζονται τα δεδομένα για να καταλήξουν στην αξιολόγηση του αλγορίθμου με τον υπολογισμό των μετρικών ακρίβειας φαίνονται στην Εικόνα 4.7. Αρχικά, η διαδικασία ξεκινάει με τα σετ δεδομένων να μετατρέπονται στην πρότυπη μορφή δεδομένων του NILMTK η οποία βασίζεται στο πρότυπο HDF5. Η ιεράρχηση των δεδομένων ενός αρχείου τέτοιου προτύπου φαίνεται παρακάτω:

```
dataset
|--- house_1
|   |--- ambient
|   |--- external
|   |--- metadata.json
|   |--- utility
|       |--- electricity
|           |--- appliances
|               |--- fridge.csv
|               |--- pump.csv
|               |--- circuits
|               |--- panel_1.csv
|               |--- mains
|                   |--- mains_1.csv
|                   |--- wiring.json
|               |--- gas
|               |--- water
|--- house_2
```

Στην συνέχεια ο χρήστης καλεί συναρτήσεις που έχουν να κάνουν με στατιστικά στοιχεία του συνόλου δεδομένων για τον εντοπισμό ζητημάτων ή για μια εξάγει στατιστική πληροφορία. Έπειτα, η εργαλειοθήκη διαθέτει συναρτήσεις που αφορούν λειτουργίες προεπεξεργασίας όπως είναι ο καθαρισμός ατελειών ή η νέα δειγματοληψία δεδομένων από τα ήδη υπάρχοντα με μικρότερη συχνότητα προκειμένου να μειωθεί ο όγκος των δεδομένων για την κύρια επεξεργασία. Ακόμα είναι κοινή πρακτική η

κανονικοποίηση των δεδομένων πριν την διαδικασία της εκπαίδευσης. Τα σύνολα δεδομένων που συλλέγονται από διαφορετικές χώρες, παρουσιάζουν διακυμάνσεις τάσης που ποικίλουν σημαντικά. Τέτοιες διακυμάνσεις μπορούν να επηρεάσουν σημαντικά τα πειράματα. Έτσι το NILMTK παρέχει συναρτήσεις κανονικοποίησης με βάση την εξίσωση του Hart:

$$Power_{normalised} = \left(\frac{Voltage_{nominal}}{Voltage_{observed}} \right)^2 \times Power_{observed}$$

Το NILMTK παρέχει έναν αριθμό υλοποιημένων αλγορίθμων που μπορούν να εκπαιδεύονται σε πραγματικά επιμερισμένα δεδομένα και στην συνέχεια να οδηγήσουν στην αποσύνθεση ολόκληρου του συνόλου των δεδομένων. Τέλος, ένα σύνολο συναρτήσεων αξιολόγησης δίνει την δυνατότητα στο να μετρηθεί η απόδοση και η αποτελεσματικότητα της αποσύνθεσης. Οι περιγραφές αυτών των μετρικών ακρίβειας δίνονται παρακάτω:

- Error in total energy assigned: Η διαφορά μεταξύ της συνολικής προβλεπόμενης ενέργειας και της πραγματικής ενέργειας που καταναλώνεται από την συσκευή η σε ολόκληρο το σύνολο δεδομένων.

$$\left| \sum_t y_t^{(n)} - \sum_t \hat{y}_t^{(n)} \right|$$

- Fraction of total energy assigned correctly: Η διαφορά μεταξύ του κλάσματος της προβλεπόμενης ενέργειας που καταναλώνεται από κάθε συσκευή και του κλάσματος πραγματικής της ενέργειας που καταναλώνεται από κάθε συσκευή προς το σύνολο δεδομένων.

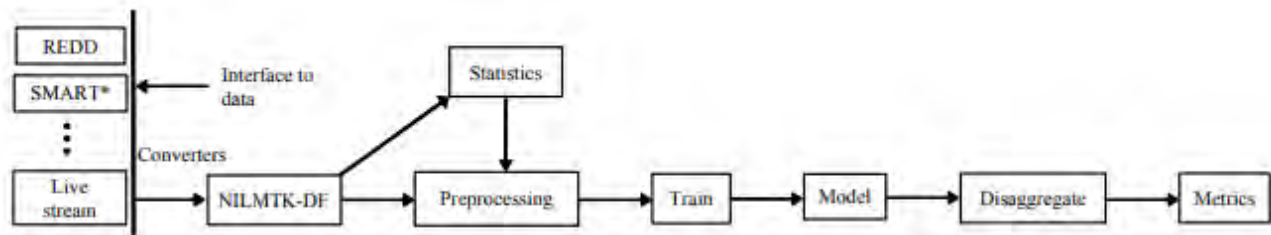
$$\sum_n \min \left(\frac{\sum_n y_t^{(n)}}{\sum_{n,t} y_t^{(n)}}, \frac{\sum_n \hat{y}_t^{(n)}}{\sum_{n,t} \hat{y}_t^{(n)}} \right)$$

- Normalised error in assigned power: Το άθροισμα των διαφορών μεταξύ της προβλεπόμενης ενέργειας και της πραγματικής ενέργειας μιας συσκευής η για κάθε χρονική στιγμή t, κανονικοποιημένο με την συνολική κατανάλωση της συσκευής.

$$\frac{\sum_t |y_t^{(n)} - \hat{y}_t^{(n)}|}{\sum_t y_t^{(n)}}$$

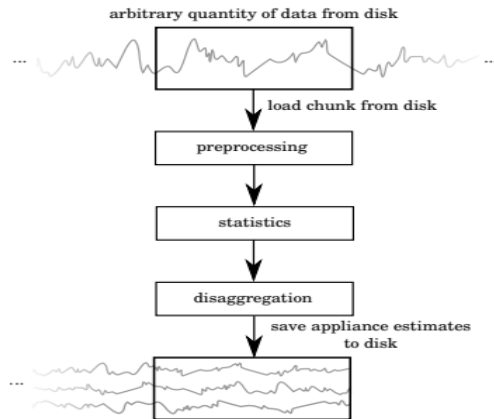
- RMS error in assigned power: Η τετραγωνική ρίζα του μέσου σφάλματος μεταξύ της της προβλεπόμενης ισχύς και της πραγματικής ισχύς της συσκευής η για κάθε χρονική στιγμή t.

$$\sqrt{\frac{1}{T} \sum_t \left(y_t^{(n)} - \hat{y}_t^{(n)} \right)^2}$$



Εικόνα 4.7: Η διαδικασία του NILMTK

Ένα από τα πλεονεκτήματα του NILMTK είναι ότι μπορεί να διαχειριστεί πολύ μεγάλο όγκο δεδομένων. Αυτό γίνεται με την φόρτωση από τον δίσκο σε κομμάτια (chunks) έτσι ώστε κάθε κομμάτι να χωρέσει στην μνήμη (Εικόνα 4.8). Η επεξεργασία εκτός πυρήνα επιτρέπει στο NILMTK να χειρίζεται σύνολα δεδομένων που είναι πολύ μεγάλα για να χωρέσουν στην μνήμη του συστήματος και επιτρέπει επίσης στον περιορισμό της χρήσης μνήμης [28]. Η διαδικασία του NILMTK φαίνεται στην εικόνα 4.7.



Εικόνα 4.8: Η φόρτωση ενός συνόλου δεδομένων από το δίσκο και

στη συνέχεια διοχετεύεται αυτό το κομμάτι μέσω ενός αγωγού επεξεργασίας.

Παρακάτω παρατίθεται μια αναπαράσταση της ροής διεργασιών του NILMTK με την χρήση κώδικα (Εικόνα 4.9).

Algorithm 1 Example code of complete pipeline.

```

dataset = DataSet()

# Load the dataset
dataset.load_hdf5(DATASET_PATH)

# Load first house
building = dataset.buildings[1]

# Remove records where voltage<160
building = filter_out_implausible_values(
    building, Measurement('voltage', ''), 160)

# Downsample to 1 minute
building = downsample(building, rule='1T')

# Choosing feature for disaggregation
DISAGG_FEATURE = Measurement('power', 'active')

# Dividing the data into train and test
train, test = train_test_split(building)

# Train on DISAGG_FEATURES using FHMM
disaggregator = FHMM()
disaggregator.train(train, disagg_features=
    [DISAGG_FEATURE])

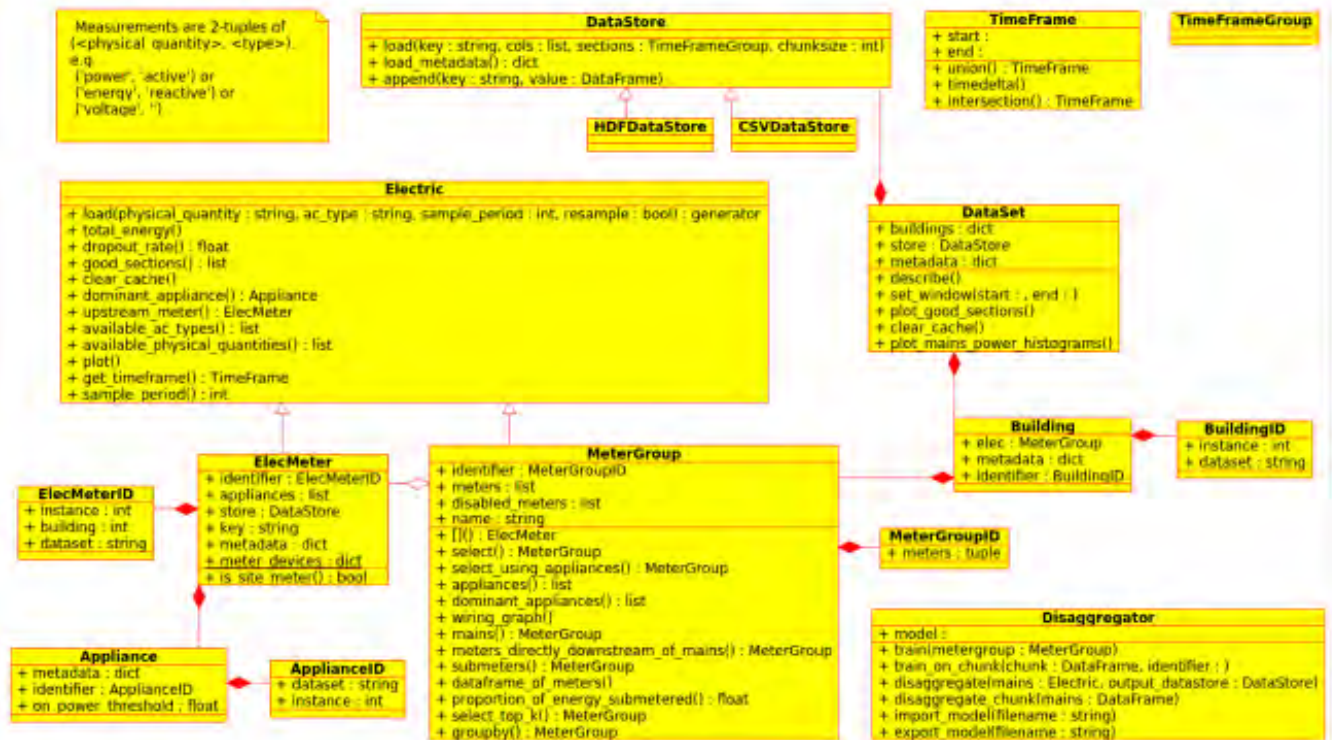
# Disaggregate
disaggregator.disaggregate(test)

# F1 score metric
f1_score = f1(disaggregator.predictions,
    test)
  
```

Εικόνα 4.9 Η ροή διεργασιών του NILMTK μέσω ενός παραδείγματος απλού κώδικα.

4.2.3 Κλάσεις

Ο σχεδιασμός του NILMTK μπορεί να θεωρηθεί ότι χωρίζεται σε στρώματα αφαίρεσης όπου το κάτω επίπεδο είναι πιο κοντά στο υλικό και το ανώτερο στρώμα παρέχει το υψηλότερο επίπεδο αφαίρεσης προκειμένου να παρέχεται μια κατάλληλη διεπαφή για τους χρήστες. Ένα UML διάγραμμα (διάγραμμα βασισμένο στην UML- Unified Modeling Language) με σκοπό την οπτική αναπαράσταση του συστήματος μαζί με τους κύριους παράγοντες, τους ρόλους, τις ενέργειες τα αντικείμενα και τις κλάσεις δίνεται παρακάτω (Εικόνα 4.10):



Εικόνα 4.10: UML διάγραμμα για το NILMTK.

Ο αντικειμενοστραφής προγραμματισμός συνδέεται άμεσα με τα αντικείμενα και κατά συνέπεια με τις κλάσεις. Μια κλάση δεν είναι τίποτα άλλο παρά ένας ορισμός. Αυτό που ορίζει είναι το ποιες ιδιότητες (attributes) και ποιες μεθόδους (methods) θα έχει ένα

αντικείμενο. Σε αυτή την ενότητα θα γίνει περιγραφή των βασικών κλάσεων του NILMTK.

- Κλάση [DataStore](#)

Η κλάση `DataStore` ορίζει τη διεπαφή για την φόρτωση και την εγγραφή δεδομένων και βρίσκεται στο κατώτερο επίπεδο αφαίρεσης του NILMTK. Φορτώνει ένα μόνο κομμάτι δεδομένων την φορά και επιστρέφει έναν πίνακα δύο διαστάσεων μέσα στην μνήμη. Η `DataStore` δεν έχει καμία γνώση σχετικά με το τι σημαίνουν τα δεδομένα. Κάθε στήλη θα μπορούσε να αντιπροσωπεύει την τάση, το ρεύμα κτλ. Επίσης, η `DataStore` έχει υλοποιημένες δυο υποκλάσεις: την `HDFDataStore` και την `CSVDataStore` για δεδομένα τύπου HDFD και CSVD αντίστοιχα. Για να ελαχιστοποιηθεί η χρήση της μνήμης, μπορεί να χρησιμοποιηθεί η υποκλάση `DataStore.load()` με την οποία μπορούν να οριστούν ακριβώς ποιες στήλες πρέπει να φορτωθούν και μπορεί να δοθεί μια λίστα από `TimeFrames` για φόρτωση από το δίσκο.

- Κλάσεις [TimeFrame](#) και [TimeFrameGroup](#)

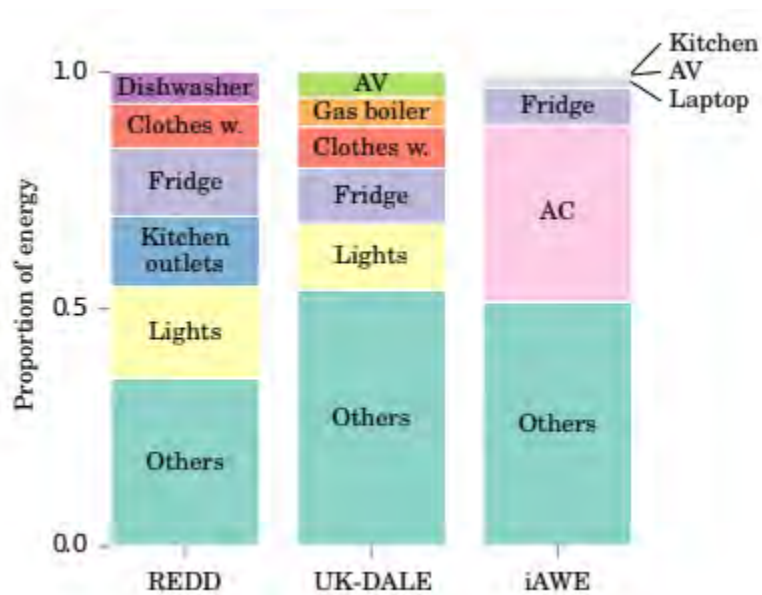
Ένα αντικείμενο `TimeFrame` αντιπροσωπεύει μια χρονική περίοδο με αυθαίρετη ημερομηνία έναρξης και μια αυθαίρετη ημερομηνία λήξης. Περιέχει μεθόδους για τον χειρισμό χρονικών περιόδων. Η κλάση `TimeFrameGroup` υπάρχει για την αποθήκευση πολλαπλών αντικειμένων `TimeFrame`. Περιέχει μεθόδους, όπως για παράδειγμα για την εύρεση της διασταύρωσης δύο αντικειμένων `TimeFrameGroup`.

- Υπερκλάση [Electric](#), κλάσεις [ElecMeter](#) και [MeterGroup](#)

Η κλάση `Electric` βρίσκεται στο αμέσως πιο πάνω επίπεδο της κλάσης `DataStore` και έχει δύο υποκλάσεις: την `ElecMeter` και την `MeterGroup`. Η κλάση `Electric` παρέχει μεθόδους όπως η `plot_power_histogram()` η οποία μπορεί να χρησιμοποιηθεί για να δημιουργήσει ιστογράμματα. Η κλάση `ElecMeter` μπορεί αντιπροσωπεύει έναν ενιαίο μετρητή ηλεκτρικής ενέργειας ενώ η κλάση `MeterGroup` πολλούς μετρητές ηλεκτρικής ενέργειας. Η `MeterGroup` περιέχει ένα σύνολο μεθόδων για την επιλογή `ElecMeters`. Για παράδειγμα, το `MeterGroup.select_top_k`

($k=5$) θα επιστρέψει ένα νέο MeterGroup που περιέχει μόνο τις πέντε πρώτες ElecMeters, ανάλογα με την κατανάλωση ενέργειας επιτρέποντας την δημιουργία γραφικών παραστάσεων (Εικόνα 4.11). Ένας άλλος τρόπος για να επιλέξουμε από ένα MeterGroup είναι να κάνουμε αναζήτηση με τον τύπο της συσκευής που είναι συνδεδεμένος στον μετρητή. Για παράδειγμα, η `metergroup['βραστήρας']` θα επιλέξει το ElecMeter που συνδέεται με τον βραστήρα.

Οι ElecMeters μπορούν να έχουν ένα σημαντικό όγκο μεταδεδομένων που σχετίζονται με αυτές.



Εικόνα 4.11: Οι πρώτες πέντε συσκευές σχετικά με την ενέργεια που καταναλώνουν που χρησιμοποιούνται σε ένα σπίτι για ορισμένα σύνολα δεδομένων.

- Κλάση [Appliance](#)

Το NILMTK διακρίνει μεταξύ συσκευών και μετρητών ηλεκτρικής ενέργειας. Η κλάση `Appliance` αντιπροσωπεύει μια μοναδική ηλεκτρική συσκευή, όπως για παράδειγμα έναν βραστήρα και η συσκευή μπορεί να έχει λεπτομερή μεταδεδομένα που σχετίζονται με αυτή. Αυτά τα μεταδεδομένα μπορούν να προέρχονται από το ίδιο το σύνολο δεδομένων (όπως η μάρκα και το μοντέλο της συσκευής) ή μπορεί να προέρχονται από το κεντρική αποθήκη μεταδεδομένων του NILM.

- Κλάσεις [Building](#) και [DataSet](#)

Η κλάση `Building` αντιπροσωπεύει ένα φυσικό κτίριο όπως ένα σπίτι. Κάθε αντικείμενο κτιρίου έχει ένα χαρακτηριστικό `elec` που είναι ένα αντικείμενο `MeterGroup` που αντιπροσωπεύει όλα τα `ElecMeters` σε αυτό το κτήριο. Κάθε αντικείμενο `Building` έχει επίσης ένα χαρακτηριστικό μεταδεδομένων που είναι ένα λεξικό Python που έχει καταγεγραμμένα μεταδεδομένα σχετικά με το κτήριο. Ένα αντικείμενο `DataSet` περιέχει όλα τα κτήρια σε ένα ενιαίο σύνολο δεδομένων και παρέχει μερικές μεθόδους για υπολογισμούς στατιστικών στοιχείων για ολόκληρο το σύνολο δεδομένων όπως την `plot_good_sections` η οποία μπορεί να χρησιμοποιηθεί για δημιουργία γραφικών παραστάσεων. Ακόμα παρέχεται η δυνατότητα για υπολογισμό στατιστικών ενός εύρους του συνόλου. Η `DataSet.set_window()` επιτρέπει τον ορισμό ενός `TimeFrame` κατά το οποίο θα πραγματοποιηθούν όλοι οι επόμενοι υπολογισμοί.

- Υπερκλάση `Disaggregator`

Αυτή η κλάση παρέχει μια κοινή διεπαφή για την εκπαίδευση και τον έλεγχο των αλγορίθμων διαχωρισμού. Υπάρχουν τέσσερις υλοποιήσεις αλγορίθμων στο NILMTK υπο την μορφή κλάσεων: `CombinatorialOptimisation`, `FHMM`, `Hart85` και `Maximum Likelihood Estimation` αποτελούν υποκλάσεις της `Disaggregator` και πρέπει να εφαρμόσουν όλες τις απαιτούμενες μεθόδους. Οι αλγόριθμοι λειτουργούν με κομμάτια δεδομένων (για οικονομία πόρων μνήμης) μέσω των μεθόδων `train_on_chunk` και `disaggregate_chunk`. Οι αλγόριθμοι κατακερματισμού πρέπει να μάθουν ένα μοντέλο για το πώς οι συσκευές καταναλώνουν ενέργεια, από τα υπάρχοντα δεδομένα. Γενικά απαιτούν δεδομένα σε επίπεδο συσκευής είτε από το κτίριο στο οποίο θα γίνει η αποσύνθεση (μάθηση υπό επίβλεψη) είτε από διαφορετικά κτίρια (μη επιτηρούμενη μάθηση). Η μέθοδος `train()` είναι η μέθοδος που καλείται για την εκπαίδευση του συστήματος και αναμένει ένα αντικείμενο `MeterGroup` (μετρητής κατανάλωσης) που θα περαστεί ως παράμετρος και που περιέχει μια λίστα αντικειμένων `ElecMeter` (μετρητής της συσκευής). Φορτώνει ένα μέρος του συνόλου των δεδομένων (`chunk`) και το μεταφέρει στην μέθοδο `train_on_chunk`. Η `train_on_chunk` καλείται ώστε τα δεδομένα τα δεδομένα των μετρητών κατανάλωσης και συσκευής να είναι παράλληλα, δηλαδή τα κομμάτια των χρονοσειρών τους να αναφέρονται στα ίδια χρονικά διαστήματα. Οι μέθοδοι

`export_model` και `import_model` είναι υπεύθυνες για την εξαγωγή από τον δίσκο και την φόρτωση ενός εκπαιδευμένου μοντέλου σε ένα αρχείο μορφής h5. Έτσι αποθηκεύεται το μοντέλο για μελλοντική χρήση ενώ διαφορετικά θα χανόταν.

- Κλάση [Metrics](#)

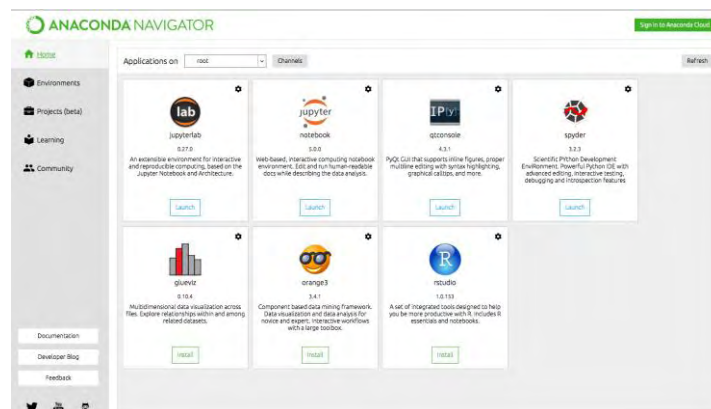
Πρόκειται για την κλάση που έρχεται τελευταία σε κάθε πείραμα αποσύνθεσης και περιέχει μεθόδους για τον υπολογισμό των σφαλμάτων και την αξιολόγηση των αλγόριθμων αποσύνθεσης.

ΚΕΦΑΛΑΙΟ 5

ΠΕΙΡΑΜΑΤΑ

5.1 Εργαλεία

Σε αυτό το κεφάλαιο θα προχωρήσουμε στην εκτέλεση περαμάτων αποσύνθεσης ενέργειας με την βοήθεια της εργαλειοθήκης NILMtk και των όσων αναφέρθηκαν παραπάνω. Πριν περάσουμε στην πράξη, θα ήταν σωστό να αναφέρουμε ότι για την εγκατάσταση και την χρήση του NILMtk, χρησιμοποιήθηκε το λογισμικό Anaconda (Εικόνα 4.12) σε περιβάλλον Windows 10, το οποίο παρέχει τα περισσότερα από τα απαιτούμενα πακέτα σε γλώσσα Python 3.6+ για την εκτέλεση πειραμάτων. Το Anaconda είναι μια πλατφόρμα ανοιχτού κώδικα, κατάλληλη για επεξεργασία δεδομένων και μηχανική μάθηση. Είναι ένα εργαλείο με το οποίο ο χρήστης μπορεί να κατεβάσει γρήγορα πακέτα σχετικά με την επιστήμη των δεδομένων, να αναπτύξει και να τρέξει αλγόριθμους μηχανικής μάθησης, να αναλύσει δεδομένα και να αναπαραστήσει τα αποτελέσματα. Κατόπιν της εγκατάστασης του παραπάνω λογισμικού, δημιουργήθηκε ένα περιβάλλον με ονομασία nilmktk-env στο οποίο εγκαταστάθηκαν τα πακέτα του NILMtk. Η ανάπτυξη και η χρήση κώδικα έγινε στο Jupiter, ένα ολοκληρωμένο διαδραστικό περιβάλλον ανάπτυξης (IDLE) που παρέχει το Anaconda.



Εικόνα 4.12: Μερικές εργαλειοθήκες και IDLE που παρέχει το Anaconda.

5.2 Εκτέλεση

5.2.1 CO-FHMM

Κάθε πείραμα αλγορίθμου μηχανικής μάθησης ξεκινάει με εκπαίδευση του μοντέλου με κατάλληλα δεδομένα στην κατάλληλη μορφή. Το NILMtk όπως έχουμε αναφέρει σε προηγούμενο κεφάλαιο, χρησιμοποιεί έναν τύπο δεδομένων βασισμένο στο πρότυπο HDF5 και ένα αρχείο τέτοιου τύπου αποθηκεύει δεδομένα μετρήσεων ενέργειας και μεταδεδομένα. Συνεπώς το πρώτο βήμα όταν χρησιμοποιούμε την εργαλειοθήκη NILMtk είναι η μετατροπή του συνόλου δεδομένων που έχουμε προμηθευτεί σε μορφή HDF5. Σε αυτό το πείραμα χρησιμοποιήθηκε το σύνολο δεδομένων REDD και η μετατροπή του έγινε όπως φαίνεται παρακάτω:

```
In [1]: from nilmtk.dataset_converters import convert_redd
        convert_redd(r'C:\Users\damaskos\Desktop\Thesis\data\REDD', 'C:\Users\
```

Στη συνέχεια το νέο αρχείο δεδομένων ορίσθηκε ως το αρχείο εισόδου του NILMTK και τυπώθηκαν τα μεταδεδομένα τα οποία φέρει μαζί με κάποια στοιχεία του:

```
In [3]: from nilmtk import DataSet
        from nilmtk.utils import print_dict

        redd = DataSet(r'C:\Users\konda\Desktop\Thesis\data\redd.h5')
```

```
In [4]: print_dict(redd.metadata)
```


- name: REDD
- long_name: The Reference Energy Disaggregation Data set
- creators:
 - Kolter, Zico
 - Johnson, Matthew
- publication_date: 2011
- institution: Massachusetts Institute of Technology (MIT)
- contact: zkoller@cs.cmu.edu
- description: Several weeks of power data for 6 different homes.
- subject: Disaggregated power demand from domestic buildings.
- number_of_buildings: 6
- timezone: US/Eastern
- geo_location:
 - locality: Massachusetts
 - country: US
 - latitude: 42.360091
 - longitude: -71.09416
- related_documents:
 - <http://redd.csail.mit.edu>
 - J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011. <http://redd.csail.mit.edu/koller-kddsust11.pdf>
- schema: https://github.com/nilmtn/nilm_metadata/tree/v0.2
- meter_devices:
 - eMonitor:
 - model: eMonitor
 - manufacturer: Powerhouse Dynamics
 - manufacturer_url: <http://powerhousedynamics.com>
 - description: Measures circuit-level power demand. Comes with 24 CTs. This FAQ page suggests the eMonitor measures real (active) power: <http://www.energycircle.com/node/14103> although the REDD readme.txt says all channels record apparent power.
 - sample_period: 3
 - max_sample_period: 50
 - measurements:
 - {physical_quantity: 'power', 'type': 'active', 'upper_limit': 5000, 'lower_limit': 0}
 - wireless: False
 - REDD_whole_house:
 - description: REDD's DIY power meter used to measure whole-home AC waveforms at high frequency. To quote from their paper: "CTs from TED (<http://www.theenergydetective.com>) to measure current in the power mains, a Pico TA041 oscilloscope probe (<http://www.picotechnologies.com>) to measure voltage for one of the two phases in the home, and a National Instruments NI-9239 analog to digital converter to transform both these analog signals to digital readings. This A/D converter has 24 bit resolution with noise of approximately 70 μ V, which determines the noise level of our current and voltage readings: the TED CTs are rated for 200 amp circuits and a maximum of 3 volts, so we are able to differentiate between currents of approximately $((200)/(70 \times 10^{-6}))(3) = 4.66$ mA, corresponding to power changes of about 0.5 watts. Similarly, since we use a 1:100 voltage stepdown in the oscilloscope probe, we can detect voltage differences of about 7mV."
 - sample_period: 1
 - max_sample_period: 30
 - measurements:
 - {physical_quantity: 'power', 'type': 'apparent', 'upper_limit': 50000, 'lower_limit': 0}
 - wireless: False


```
In [5]: print_dict(redd.buildings)
```

- 1: Building(instance=1, dataset='REDD')
- 2: Building(instance=2, dataset='REDD')
- 3: Building(instance=3, dataset='REDD')
- 4: Building(instance=4, dataset='REDD')
- 5: Building(instance=5, dataset='REDD')
- 6: Building(instance=6, dataset='REDD')

```
In [6]: print_dict(redd.buildings[1].metadata)
```

- **instance:** 1
- **original_name:** house_1
- **dataset:** REDD

Με τις παρακάτω εντολές βλέπουμε ποιες μετρήσεις έχουν παρθεί για μια συγκεκριμένη συσκευή και τις παρουσιάζουμε σε έναν πίνακα:

```
In [10]: elec=redd.buildings[1].elec
```

```
In [11]: fridge = elec['fridge']
fridge.available_columns()
```

```
Out[11]: [('power', 'active')]
```

```
In [12]: df = next(fridge.load())
df.head()
```

```
Out[12]:
```

physical_quantity	power
type	active
2011-04-18 09:22:13-04:00	6.0
2011-04-18 09:22:16-04:00	6.0
2011-04-18 09:22:20-04:00	6.0
2011-04-18 09:22:23-04:00	6.0
2011-04-18 09:22:26-04:00	6.0

Κάθε σύνολο μετρήσεων περιέχει ποικίλες μετρήσεις ενέργειας. Το NILMtk μας παρέχει την δυνατότητα να συλλέξουμε, να τις χωρίσουμε σε διάφορα υποσύνολα και να τις ταξινομήσουμε με βάση κάποια κριτήρια όπως για παράδειγμα το είδος της συσκευής ή η κατανάλωση ενέργειας.

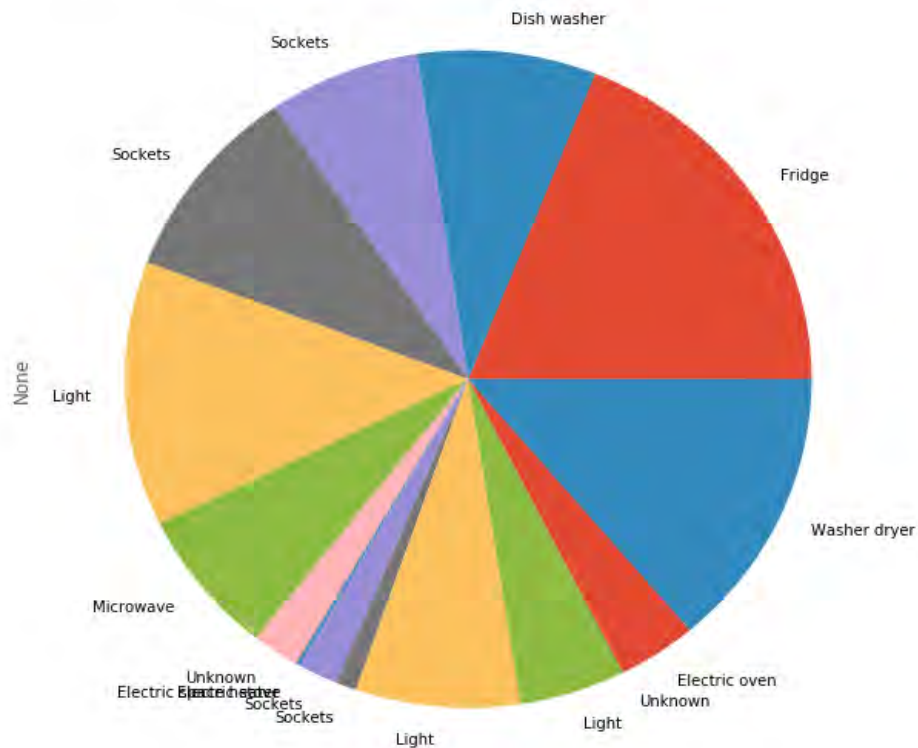
Παραδείγματος χάρη, με τις παρακάτω εντολές δημιουργούμε ένα υποσύνολο που περιέχει μόνο τις μετρήσεις οι οποίες είχαν ισχύ μεγαλύτερη από 20kWh.

```
In [14]: energy_per_meter = energy_per_meter.loc['active']
more_than_20 = energy_per_meter[energy_per_meter > 20]
more_than_20
```

```
Out[14]: (5, 1, REDD)          44.750925
(8, 1, REDD)          22.939649
(9, 1, REDD)          30.734511
(((10, 1, REDD), (20, 1, REDD)),)  32.614809
Name: active, dtype: float64
```

Ακόμα το NILMtk παρέχει εργαλεία δημιουργίας γραφικών παραστάσεων. Με τις παρακάτω συναρτήσεις γίνεται αναπαράσταση της συνολικής ενέργειας που καταναλώνει κάθε συσκευή σε σχέση με τις υπόλοιπες αλλά και η γραφική παράσταση μιας συγκεκριμένης σε συνάρτηση με τον χρόνο.

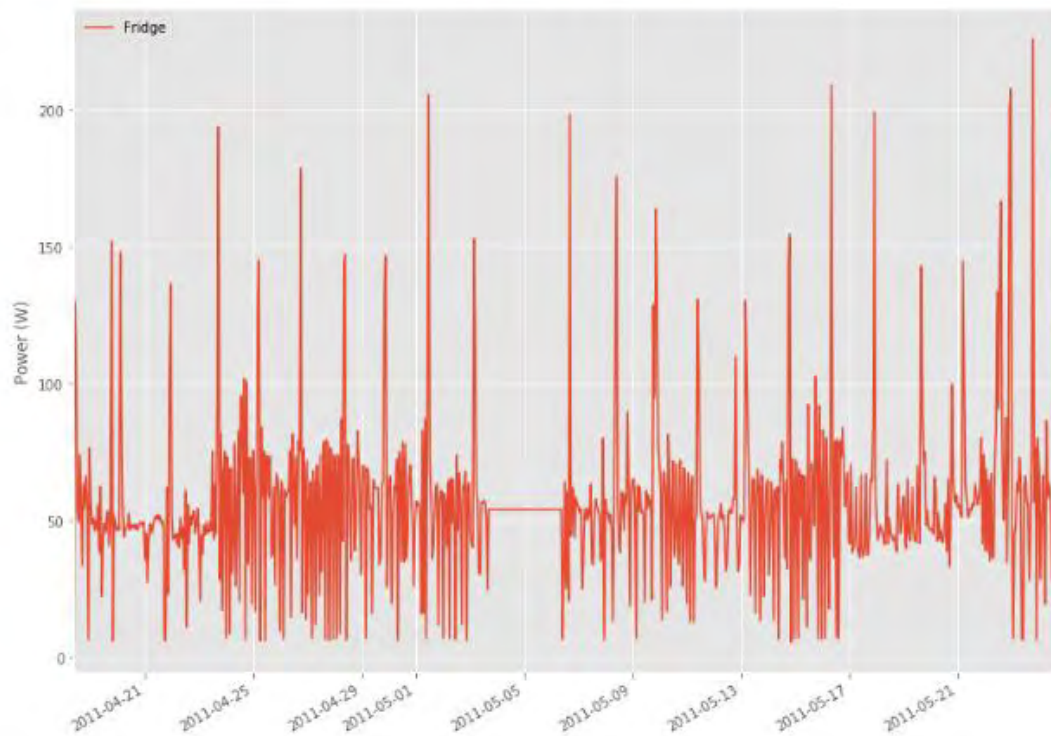
```
In [18]: labels = elec.get_labels(fraction.index)
plt.figure(figsize=(10,30))
fraction.plot(kind='pie', labels=labels);
```



```
In [20]: fridge_meter = elec['fridge']
```

```
In [21]: fridge_meter.plot()
```

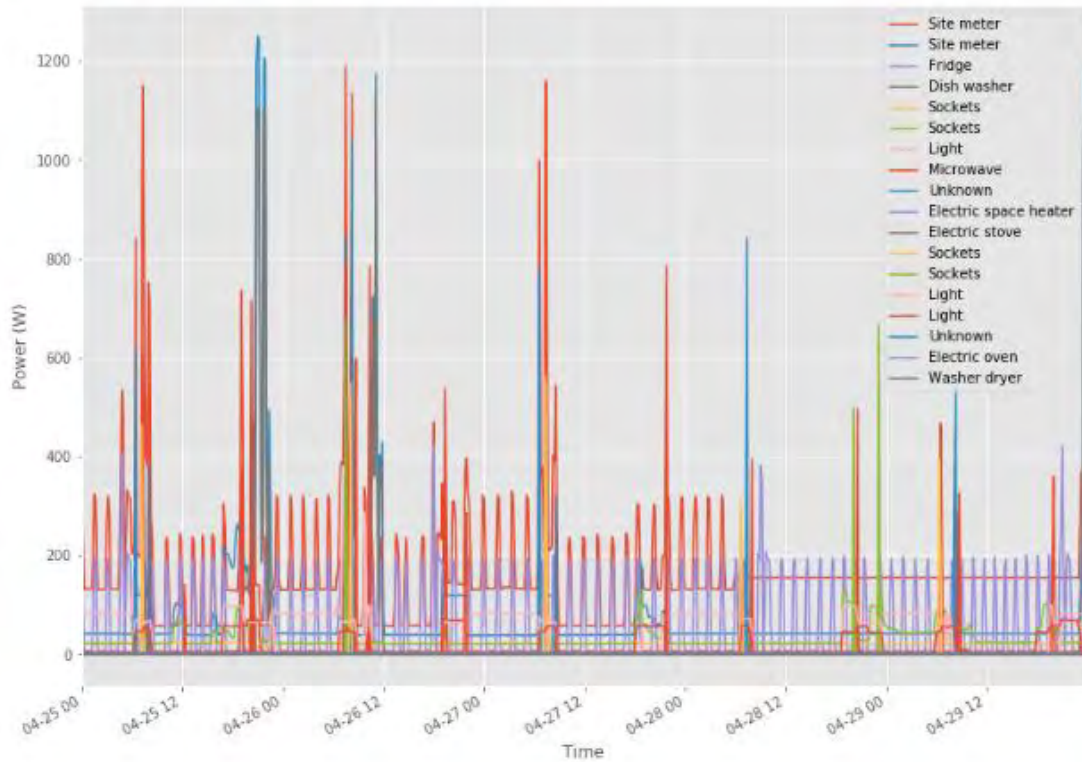
```
Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x266488b1a90>
```



Ακόμα μπορούμε να θέσουμε ένα ορισμένο χρονικό διάστημα και να αναπαραστήσουμε τις μετρήσεις όλων των συσκευών για εκείνο:

```
In [22]: redd.set_window(start='2011-04-25', end='2011-04-30')
elec.plot();
plt.xlabel("Time");
```

Loading data for meter ElecMeterID(instance=4, building=1, dataset='REDD')
Done loading data all meters for this chunk.
Loading data for meter ElecMeterID(instance=20, building=1, dataset='REDD')
Done loading data all meters for this chunk.



Περνάμε τώρα στο κομμάτι της εκπαίδευσης του μοντέλου, του επιμερισμού της ενέργειας και της αξιολόγησης των αποτελεσμάτων. Μετά την εισαγωγή των απαραίτητων κλάσεων προχωράμε στην διαίρεση του συνόλου σε ένα σετ που θα χρησιμοποιηθεί για την εκπαίδευση και ένα για την αξιολόγηση των αλγορίθμων αποσύνθεσης.

```
In [23]: from __future__ import print_function, division
import time

from matplotlib import rcParams
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from six import iteritems

from nilmtk import DataSet, TimeFrame, MeterGroup, HDFDataStore
from nilmtk.legacy.disaggregate import CombinatorialOptimisation, FHMM
import nilmtk.utils

%matplotlib inline

C:\Users\konda\Anaconda3\envs\nilmtk-env\lib\importlib\_bootstrap.py:219: RuntimeWarning: numpy.ufunc size changed, may indicate binary incompatibility. Expected 192 from C header, got 216 from PyObject
    return f(*args, **kwargs)
```

```
In [26]: train = DataSet(r'C:\Users\konda\Desktop\Thesis\data\redd.h5')
test = DataSet(r'C:\Users\konda\Desktop\Thesis\data\redd.h5')
```

Επιλέγουμε την πρώτη κατοικία και στην συνέχεια τις πρώτες πέντε ηλεκτρικές συσκευές της σε κατανάλωση:

```
In [27]: building = 1
```

```
In [28]: train.set_window(end="2011-04-20")
test.set_window(start="2011-04-20")

train_elec = train.buildings[1].elec
test_elec = test.buildings[1].elec
```

```
In [29]: top_5_train_elec = train_elec.submeters().select_top_k(k=5)

15/16 MeterGroup(meters==19, building=1, dataset='REDD', appliances=[Appliance
(type='unknown', instance=2)])e=1)])ce=1)])
    ElecMeter(instance=3, building=1, dataset='REDD', appliances=[Appliance(type
='electric oven', instance=1)])
    ElecMeter(instance=4, building=1, dataset='REDD', appliances=[Appliance(type
='electric oven', instance=1)])
16/16 MeterGroup(meters= for ElecMeterID(instance=4, building=1, dataset='RED
D') ...
    ElecMeter(instance=10, building=1, dataset='REDD', appliances=[Appliance(type
='washer dryer', instance=1)])
    ElecMeter(instance=20, building=1, dataset='REDD', appliances=[Appliance(type
='washer dryer', instance=1)])
Calculating total_energy for ElecMeterID(instance=10, building=1, dataset='RED
D') ...    total_energy for ElecMeterID(instance=20, building=1, dataset='RED
D') ...
```


Έπειτα με την βοήθεια μιας συνάρτησης αποσύνθεσης προχωράμε στην μάθηση και τον επιμερισμό της συνολικής ενέργειας στις παραπάνω πέντε συσκευές με τους αλγορίθμους CO και FHMM για το χρονικό διάστημα που έχουμε ορίσει:

```
In [32]: classifiers = {'CO':CombinatorialOptimisation(), 'FHMM':FHMM()}
predictions = {}
sample_period = 120
for clf_name, clf in classifiers.items():
    print("*"*20)
    print(clf_name)
    print("*"*20)
    start = time.time()
    # Note that we have given the sample period to downsample the data to 1 minute.
    # If instead of top_5 we wanted to train on all appliance, we would write
    # fhmm.train(train_elec, sample_period=60)
    clf.train(top_5_train_elec, sample_period=sample_period)
    end = time.time()
    print("Runtime =", end-start, "seconds.")
    gt, predictions[clf_name] = predict(clf, test_elec, sample_period, train.metadata['timezone'])
```

```
*****
CO
*****
Training model for submeter 'ElecMeter(instance=5, building=1, dataset='REDD',
appliances=[Appliance(type='fridge', instance=1)])'
Training model for submeter 'ElecMeter(instance=9, building=1, dataset='REDD',
appliances=[Appliance(type='light', instance=1)])'
Training model for submeter 'ElecMeter(instance=17, building=1, dataset='REDD',
appliances=[Appliance(type='light', instance=2)])'
Training model for submeter 'ElecMeter(instance=6, building=1, dataset='REDD',
appliances=[Appliance(type='dish washer', instance=1)])'
Training model for submeter 'ElecMeter(instance=8, building=1, dataset='REDD',
appliances=[Appliance(type='sockets', instance=2)])'
Done training!
Runtime = 0.8387303352355957 seconds.
Loading data for meter ElecMeterID(instance=2, building=1, dataset='REDD')
Done loading data all meters for this chunk.
Estimating power demand for 'ElecMeter(instance=5, building=1, dataset='REDD',
appliances=[Appliance(type='fridge', instance=1)])'
Estimating power demand for 'ElecMeter(instance=9, building=1, dataset='REDD',
appliances=[Appliance(type='light', instance=1)])'
Estimating power demand for 'ElecMeter(instance=17, building=1, dataset='REDD',
appliances=[Appliance(type='light', instance=2)])'
Estimating power demand for 'ElecMeter(instance=6, building=1, dataset='REDD',
appliances=[Appliance(type='dish washer', instance=1)])'
Estimating power demand for 'ElecMeter(instance=8, building=1, dataset='REDD',
appliances=[Appliance(type='sockets', instance=2)])'
Loading data for meter ElecMeterID(instance=4, building=1, dataset='REDD')
Done loading data all meters for this chunk.
Loading data for meter ElecMeterID(instance=20, building=1, dataset='REDD')
Done loading data all meters for this chunk.
*****
FHMM
*****
Training model for submeter 'ElecMeter(instance=5, building=1, dataset='REDD',
appliances=[Appliance(type='fridge', instance=1)])' with 3 states
Training model for submeter 'ElecMeter(instance=9, building=1, dataset='REDD',
appliances=[Appliance(type='light', instance=1)])' with 3 states
Training model for submeter 'ElecMeter(instance=17, building=1, dataset='REDD',
appliances=[Appliance(type='light', instance=2)])' with 3 states
Training model for submeter 'ElecMeter(instance=6, building=1, dataset='REDD',
appliances=[Appliance(type='dish washer', instance=1)])' with 3 states
Training model for submeter 'ElecMeter(instance=8, building=1, dataset='REDD',
appliances=[Appliance(type='sockets', instance=2)])' with 3 states
Runtime = 4.831832408905029 seconds.
Loading data for meter ElecMeterID(instance=2, building=1, dataset='REDD')
Done loading data all meters for this chunk.
```

Παρακάτω βλέπουμε τα αποτελέσματα των αλγορίθμων:

```
In [33]: appliance_labels = [m.label() for m in gt.columns.values]
```

```
In [34]: gt.columns = appliance_labels  
predictions['CO'].columns = appliance_labels  
predictions['FHMM'].columns = appliance_labels
```

```
In [35]: gt.head()
```

Out[35]:

	Fridge	Light	Light	Dish washer	Sockets
2011-04-20 00:00:00-04:00	6.225806	81.096771	0.0	0.000000	22.000000
2011-04-20 00:02:00-04:00	6.281250	81.218750	0.0	0.000000	22.031250
2011-04-20 00:04:00-04:00	67.718750	81.562500	0.0	0.125000	21.843750
2011-04-20 00:06:00-04:00	204.161285	82.000000	0.0	0.193548	21.064516
2011-04-20 00:08:00-04:00	196.125000	82.031250	0.0	0.125000	21.250000

```
In [36]: predictions['CO'].head()
```

Out[36]:

	Fridge	Light	Light	Dish washer	Sockets
2011-04-20 00:00:00-04:00	0.0	82.0	64.0	0.0	24.0
2011-04-20 00:02:00-04:00	0.0	82.0	96.0	0.0	0.0
2011-04-20 00:04:00-04:00	58.0	0.0	96.0	0.0	67.0
2011-04-20 00:06:00-04:00	58.0	0.0	64.0	227.0	24.0
2011-04-20 00:08:00-04:00	58.0	278.0	0.0	0.0	24.0

```
In [37]: predictions['FHMM'].head()
```

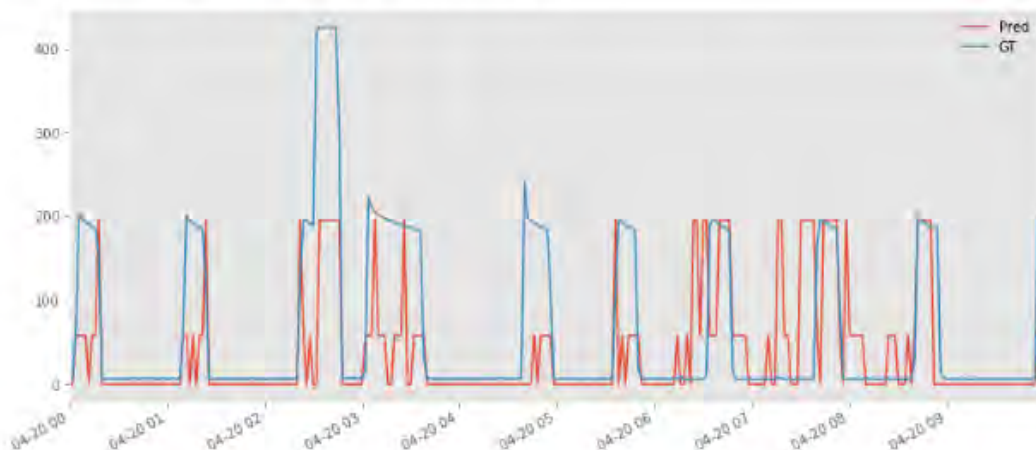
Out[37]:

	Fridge	Light	Light	Dish washer	Sockets
2011-04-20 00:00:00-04:00	6.0	81.0	64.0	0.0	22.0
2011-04-20 00:02:00-04:00	6.0	81.0	64.0	0.0	22.0
2011-04-20 00:04:00-04:00	6.0	81.0	61.0	61.0	22.0
2011-04-20 00:06:00-04:00	146.0	81.0	61.0	61.0	22.0
2011-04-20 00:08:00-04:00	146.0	192.0	0.0	0.0	22.0

Επίσης οι γραφικές παραστάσεις των αποτελεσμάτων για μια συσκευή και η σύγκριση με τις πραγματικές μετρήσεις της:

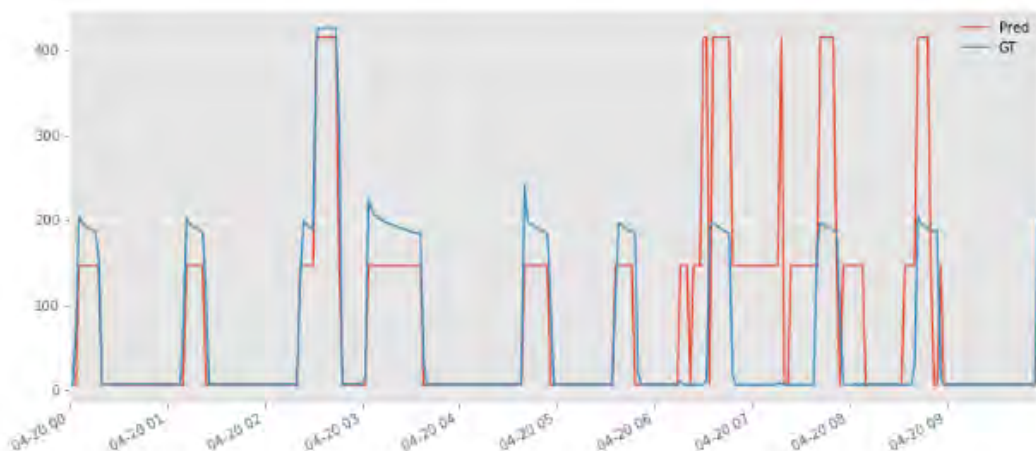
```
In [38]: predictions['CO']['Fridge'].head(300).plot(label="Pred")
gt['Fridge'].head(300).plot(label="GT")
plt.legend()
```

Out[38]: <matplotlib.legend.Legend at 0x2664dcd3780>



```
In [39]: predictions['FHMM']['Fridge'].head(300).plot(label="Pred")
gt['Fridge'].head(300).plot(label="GT")
plt.legend()
```

Out[39]: <matplotlib.legend.Legend at 0x2664dc09fd0>



Τελικά, περνάμε στον υπολογισμό μετρικών ακρίβειας και συγκεκριμένα τον υπολογισμό της τιμής RMSE (Root Mean Square Error) που είναι ένας τρόπος υπολογισμού του βαθμού σφάλματος που είχε το μοντέλο για κάθε αλγόριθμο:


```
In [41]: rmse = {}
for clf_name in classifiers.keys():
    rmse[clf_name] = nilmtk.utils.compute_rmse(gt, predictions[clf_name])

rmse = pd.DataFrame(rmse)
rmse
```

Out[41]:

	CO	FHMM
Fridge	100.070580	162.850996
Light	63.869968	47.356238
Dish washer	404.663116	390.239409
Sockets	27.551142	30.514786

Όμοια για το σύνολο δεδομένων iawe υπολογίζουμε:

Out[14]:

	CO	FHMM
Air conditioner (<bound method ElecMeter.building of ElecMeter(instance=5, building=1, dataset='IAWE', appliances=[Appliance(type='air conditioner', instance=2)])>, <bound method ElecMeter.instance of ElecMeter(instance=5, building=1, dataset='IAWE', appliances=[Appliance(type='air conditioner', instance=2)])>)	1016.515564	1049.536342
Air conditioner (<bound method ElecMeter.building of ElecMeter(instance=4, building=1, dataset='IAWE', appliances=[Appliance(type='air conditioner', instance=1)])>, <bound method ElecMeter.instance of ElecMeter(instance=4, building=1, dataset='IAWE', appliances=[Appliance(type='air conditioner', instance=1)])>)	1565.822632	1480.533570
Fridge	72.456818	59.730028
Computer	39.378902	29.410414
Television	48.204884	4.310419

5.2.2 Hart

Σε αυτήν την ενότητα θα εκτελέσουμε ένα πείραμα επιμερισμού ενέργειας με την χρήση του αλγορίθμου του Hart και θα χρησιμοποιήσουμε το σύνολο δεδομένων Redd.

```

In [1]: %matplotlib inline

In [2]: import numpy as np
import pandas as pd
from pylab import rcParams
import matplotlib.pyplot as plt

rcParams['figure.figsize'] = (13, 6)

In [3]: import nilmtk
from nilmtk import DataSet, TimeFrame, MeterGroup, HDFDataStore
from nilmtk.legacy.disaggregate import CombinatorialOptimisation
from nilmtk.utils import print_dict
from nilmtk.metrics import f1_score

In [4]: data = DataSet(r'C:\Users\konda\Desktop\Thesis\data\redd.h5')
print('Loaded', len(data.buildings), 'buildings')

Loaded 6 buildings

In [5]: elec = data.buildings[2].elec

In [6]: elec.get_timeframe()

Out[6]: TimeFrame(start='2011-04-17 19:18:27-04:00', end='2011-05-22 19:59:16-04:00', empty=False)

In [7]: data.set_window(start='2011-04-20', end='2011-04-24')
elec_1 = data.buildings[2].elec

In [8]: mains=elec_1.mains()
submeters=elec_1.submeters()

In [9]: mains.available_ac_types('power')

Out[9]: ['apparent']

In [10]: submeters.available_ac_types('power')

Out[10]: ['active']

In [12]: from nilmtk.legacy.disaggregate.hart_85 import Hart85
h = Hart85()

In [13]: h.train(mains, columns=[('power', 'apparent')])

Loading data for meter ElecMeterID(instance=2, building=2, dataset='REDD')
Done loading data all meters for this chunk.
Finding Edges, please wait ...
Edge detection complete.
Creating transition frame ...
Transition frame created.
Creating states frame ...
States frame created.
Finished.

In [14]: #h.export_model('model.pickle')

In [15]: #h.import_model('model.pickle')

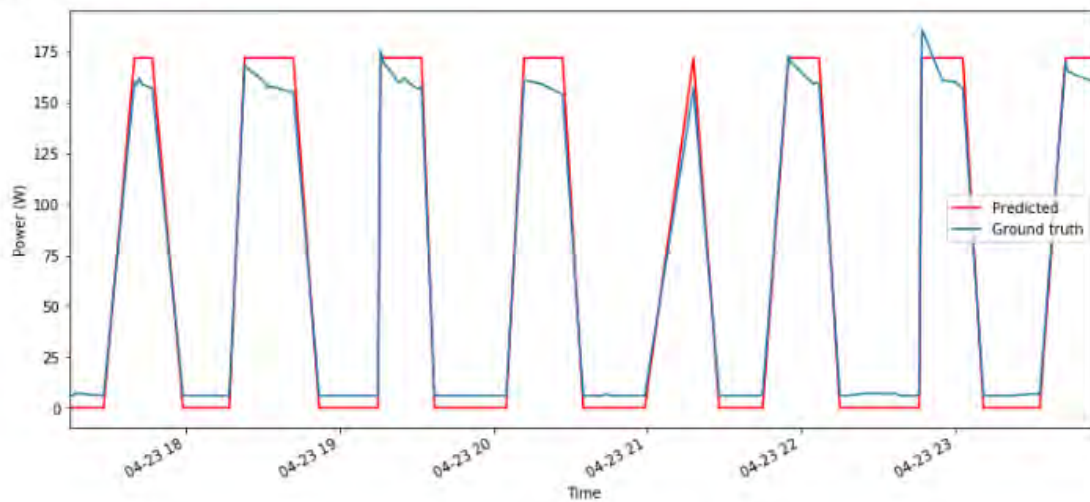
In [16]: output = HDFDataStore('output.h5', 'w')
df=h.disaggregate(mains, output)

```

Για την συσκευή ψυγείο (fridge) παρατηρούμε στην παρακάτω γραφική παράσταση ικανοποιητική πρόβλεψη:

```
df_fridge = next(elec_1['fridge', 1].load())
merged_df = pd.merge(df[0], df_fridge, left_index=True, right_index=True)
```

```
In [23]: merged_df[0].plot(c='r')
merged_df['power', 'active'].plot()
plt.legend(["Predicted", "Ground truth"]);
plt.ylabel("Power (W)")
plt.xlabel("Time");
```



ενώ για την συσκευή στεγνωτήρα (washer dryer) ο αλγόριθμος απέτυχε να προβλέψει:

```
In [19]: df_dish_washer = next(elec_1['washer dryer', 1].load())
merged_df = pd.merge(df[2], df_dish_washer, left_index=True, right_index=True)
```

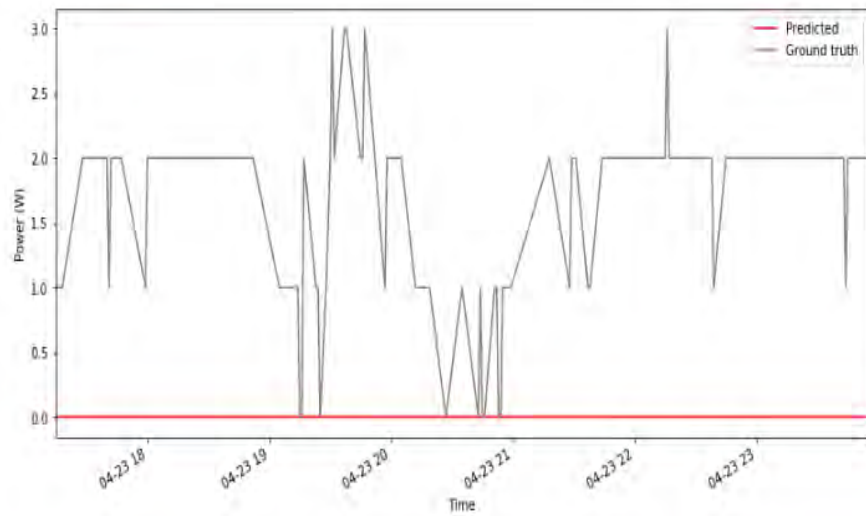
C:\Users\konda\Anaconda3\envs\nilmtk-env\lib\site-packages\pandas\core\reshape\merge.py:617: UserWarning: merging between different levels can give an unintended result (1 levels on the left, 2 on the right)
warnings.warn(msg, UserWarning)

```
In [20]: merged_df.head()
```

```
Out[20]:
```

	2	(power, active)
2011-04-23 17:15:00-04:00	0	1.0
2011-04-23 17:16:00-04:00	0	1.0
2011-04-23 17:17:00-04:00	0	1.0
2011-04-23 17:18:00-04:00	0	1.0
2011-04-23 17:28:00-04:00	0	2.0

```
In [21]: ax1 = merged_df[2].plot(c='r')
ax2 = merged_df['power', 'active'].plot(c='grey')
ax1.legend(["Predicted", "Ground truth"])
plt.ylabel("Power (W)")
plt.xlabel("Time");
```



ΚΕΦΑΛΑΙΟ 6

ΕΠΙΛΟΓΟΣ

6.1 Δυσκολίες

Η εργαλειοθήκη NILMitk αναπτύχθηκε με σκοπό την διευκόλυνση της εκτέλεσης πειραμάτων επιμερισμού ενέργειας και την αξιολόγηση και σύγκριση των αποτελεσμάτων. Ωστόσο, η υλοποίηση του δεν έχει ακόμα ολοκληρωθεί καθώς βρίσκεται σε ερευνητικό στάδιο και πολλές συναρτήσεις του δεν λειτουργούν. Η λειτουργία του είναι αρκετά περίπλοκη για έναν αρχάριο στον χώρο αυτό και οι διαθέσιμες πηγές από όπου μπορεί να αντλήσει κάποιος πληροφορία σχετικά με την χρήση του είναι λιγοστές. Επίσης χρειάζεται βαθύτερες γνώσεις προγραμματισμού αφού απαιτείται επεξεργασία του κώδικα σε αρκετά σημεία.

6.2 Συμπεράσματα

Για την αξιολόγηση και την σύγκριση των αλγορίθμων CO και FHMM χρησιμοποιήθηκαν τα σύνολα δεδομένων REDD και iAWE. Οι αλγόριθμοι και στα δυο σύνολα δεδομένων παρουσίασαν σημαντικές αποκλίσεις στις προβλέψεις τους γεγονός που υποδεικνύει την αναγκαιότητα για την βελτίωσή τους. Η σύγκρισή τους δεν οδηγεί σε σαφή συμπεράσματα καθώς για ορισμένες συσκευές παρουσιάζουν παρόμοιες αποκλίσεις ενώ για άλλες μεγαλύτερες. Αξίζει να αναφερθεί η ακρίβεια της πρόβλεψης για την δεύτερη συσκευή κλιματιστικού (air conditioner) στο σύνολο δεδομένων iAWE ή οποία είναι αρκετά χειρότερη από αυτή του πρώτου κλιματιστικού. Αυτό οφείλεται στο γεγονός ότι το δεύτερο κλιματιστικό είχε επιλεγθεί να λειτουργεί σε μια μη σταθερή κατάσταση λειτουργίας, πράγμα που ανάγκαζε την αντλία θερμότητάς (που έχει την μεγαλύτερη κατανάλωση) του να διακόπτει συχνά την λειτουργία της σε αντίθεση με το πρώτο κλιματιστικό που δούλευε συνεχόμενα σε σταθερή κατάσταση. Από την άλλη πλευρά ο αλγόριθμος του Hart πρόβλεψε με αρκετή ακρίβεια την κατανάλωση της συσκευής του ψυγείου ενώ για την συσκευή στεγνωτήρα απέτυχε παταγωδώς.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Α. Καπλανή, «Το Ενεργειακό πρόβλημα στην Σύγχρονη Εποχή», 2013.
- [2] Π. Μελαχροινούδης, «Η μετάβαση στα Ευφυή Ηλεκτρικά Δίκτυα και συμβολή της απόκρισης της Ζήτησης στην αξιοπιστία των Δικτύων Διανομής», 2015.
- [3] Δ. Χρυσάφης, «Κατηγοριοποίηση φορτίων ΧΤ με βάση τις ηλεκτρονικές υπογραφές τους», 2012 .
- [4] George W. Hart, «Nonintrusive Appliance Load Monitoring», 1992.
- [5] Jack (Daniel) Kelly, «Disaggregation of Domestic Smart Meter Energy Data», 2017.
- [6] Μ. Ευαγγελία, «Επιμερισμός καταναλισκόμενης ενέργειας συσκευών σε ένα έξυπνο δίκτυο ηλεκτρικής ενέργειας (Smart Power Grid)», 2018.
- [7] Behzad Najafi, Sadaf Moaveninejad and Fabio Rinaldi, «Data analytics for energy disaggregation: methods and applications», 2018.
- [8] Ο. Κρυσταλακος, «Study of Neural Network Systems on Non-Intrusive Load Monitoring», 2017.
- [9] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imran and Sutharshan Rajasegarar , «Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey», 2012.
- [10] Christoforos Nalmpantis, Dimitris Vrakas, «Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparison», 2018.
- [11] WANG Hua, MA Cuiqin, ZHOU Lijuan , «A Brief Review of Machine Learning and its Application», 2009.
- [12] Fabio Rinaldi, «Data analytics for energy disaggregation: methods and applications», Chapter 3.5, 2018.
- [13] Ben Krose, Patrick van der Smagt, «An introduction to Neural Networks», 1996.
- [14] Yann LeCun, Yoshua Bengio & Geoffrey Hinton, «Deep learning», 2015.
- [15] The University of Ayckland, «Chapter 8: Markov Chains».

- [16] Zoubin Ghahramani, «An Introduction to Hidden Markov Models and Bayesian Networks», 2001.
- [17] Jurgen Van Gael, Yee Whye Teh, «The Infinite Factorial Hidden Markov Model», 2017.
- [18] J. Zico Kolter, Siddarth Batra, Andrew Y. Ng «Energy Disaggregation via Discriminative Sparse Coding ».
- [19] Oliver Parson, «NIALM as a combinatorial optimisation problem», 2011.
- [20] Oliver Parson, «Using a single hidden Markov model to model multiple appliances», 2011.
- [21] Κυρίτσης Κωνσταντίνος, «Αρχιτεκτονικές Υλικού ΓΙΑ Αποκωδικοποίηση VITERBI σε Ασύρματα Δίκτυα», 2013.
- [22] J. Zico Kolter, Matthew J. Johnson, « REDD: A Public Data Set for Energy Disaggregation Research », 2011.
- [23] Jack Kelly & William Knottenbelt, «The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes», 2015.
- [24] Kyle Anderson, Adrian Filip Ocneanu, Diego Benítez, «BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research», 2012.
- [25] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, «The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms», 2014.
- [26] www.NilmDatasets.com.
- [27] GitHub, NILMtk, «Writing a disaggregation algorithm».
- [28] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, Mani Srivastava « NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring », 2014.